

Method for noise suppression in a microphone signal

Publication number: EP0948237

Publication date: 1999-10-06

Inventor: THOMAS HANS-JOERG (DE)

Applicant: DAIMLERCHRYSLER AEROSPACE AG (DE)

Classification:

- **international:** G10L21/02; H04R3/00; G10L21/00; H04R3/00; (IPC1-7): H04R3/00

- **European:** G10L21/02A1; H04R3/00C

Application number: EP19990106123 19990401

Priority number(s): DE19981014971 19980403

Also published as:

US6895095 (B1)

EP0948237 (A3)

DE19981014971 (A1)

Cited documents:

EP0250048

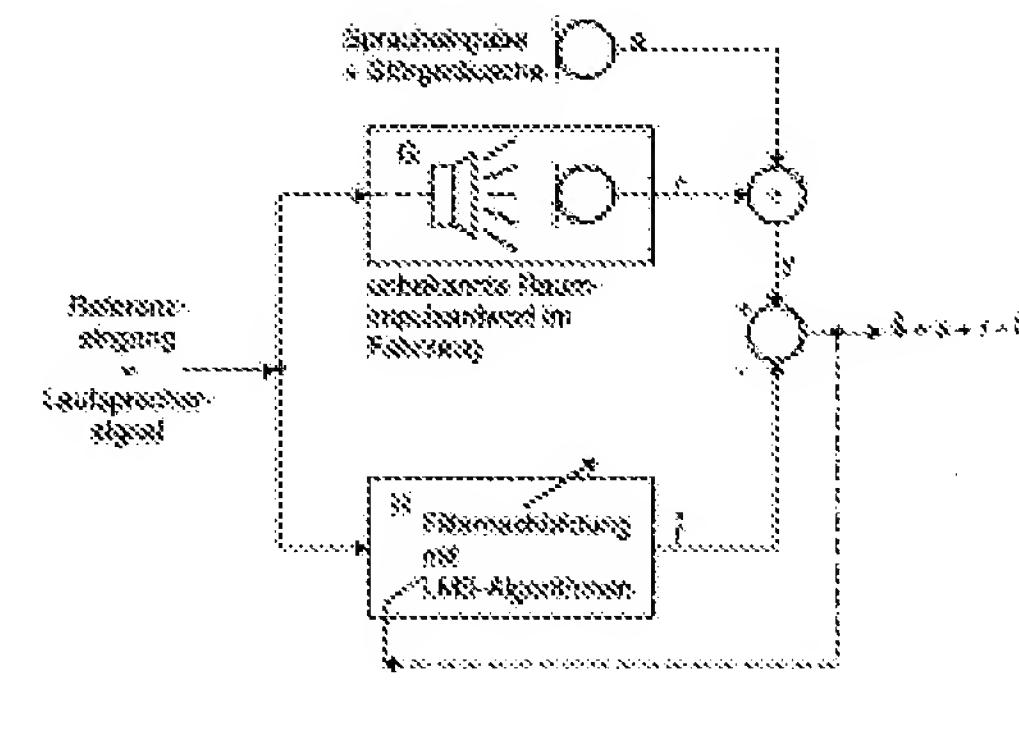
US5245664

XP004108821

[Report a data error here](#)

Abstract of EP0948237

The noise reduction method uses subtraction of a noise compensation signal from the microphone signal (y), the noise compensation signal provided by simulation of the noise signal via an adaptive filter (H) using a least mean square algorithm, supplied with a reference signal, e.g. a loudspeaker signal, with transformation of the microphone signal, the compensation signal and the output signal into the frequency range.



Data supplied from the **esp@cenet** database - Worldwide



(19) Europäisches Patentamt

European Patent Office

Office européen des brevets



(11) EP 0 948 237 A2

(12)

EUROPÄISCHE PATENTANMELDUNG

(43) Veröffentlichungstag:
06.10.1999 Patentblatt 1999/40(51) Int. Cl.⁶: H04R 3/00

(21) Anmeldenummer: 99106123.5

(22) Anmeldetag: 01.04.1999

(84) Benannte Vertragsstaaten:
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
MC NL PT SE**
Benannte Erstreckungsstaaten:
AL LT LV MK RO SI

(30) Priorität: 03.04.1998 DE 19814971

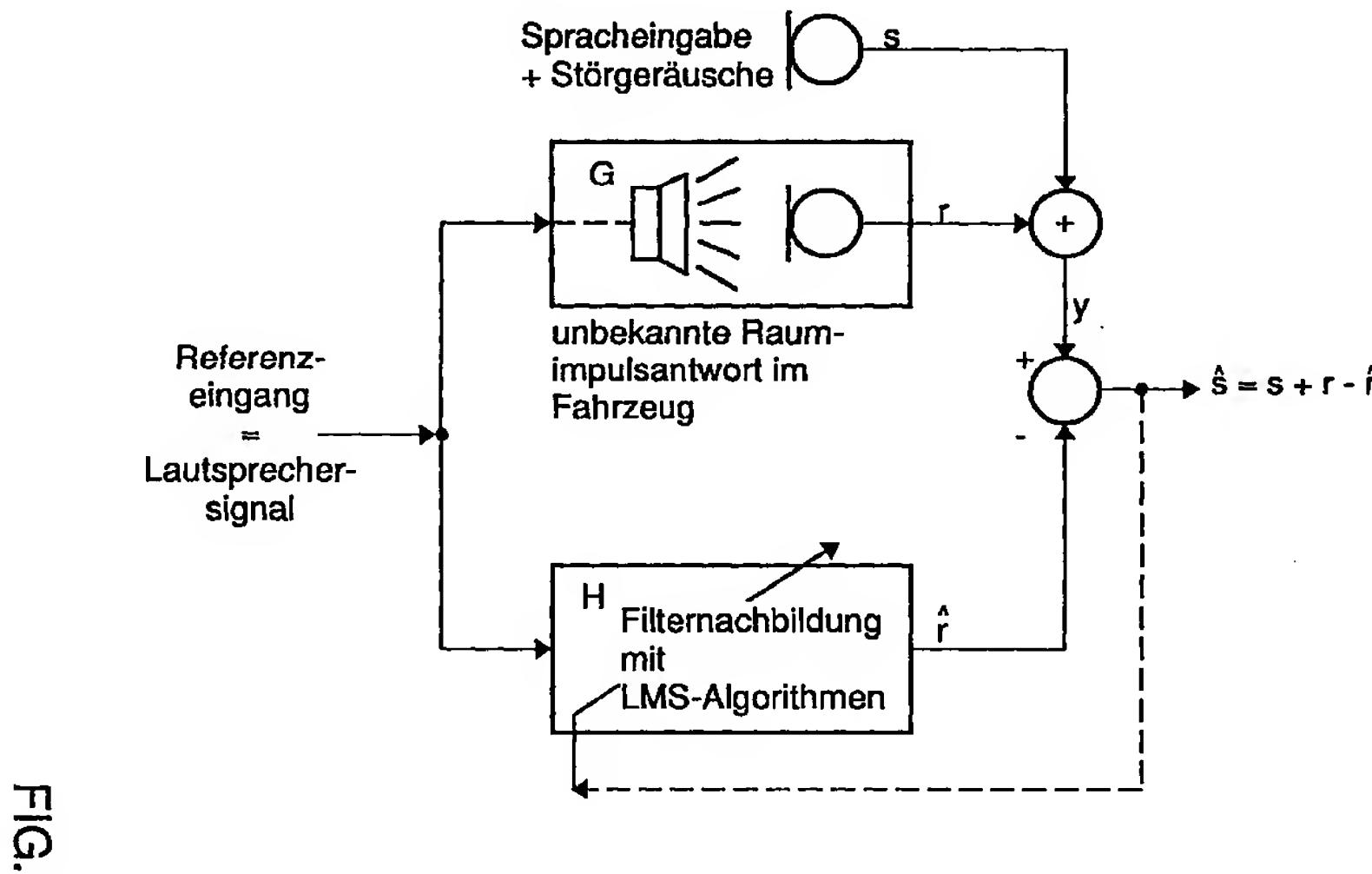
(71) Anmelder:
**DaimlerChrysler Aerospace AG
81663 München (DE)**

(72) Erfinder: **Thomas, Hans-Jörg
89233 Neu-Ulm (DE)**

(54) Verfahren zur Störbefreiung eines Mikrophonsignals

(57) Für ein Verfahren zur Befreiung eines Mikrophonsignals von Störsignalanteilen durch Erzeugen eines Kompensationssignals und Subtraktion des Kompensationssignals vom Mikrophonsignal wird vorgeschlagen, die Kompensation vollständig im Frequenzbereich vorzunehmen und auch das Ausgangssignal im Frequenzbereich zu verarbeiten. Es werden Maßnahmen zur Verringerung des Aufwands bei der Signalverarbeitung angegeben. Vorteilhafte Weiterbildungen sehen beispielsweise vor, daß zur

Störbefreiung eines Sprachsignals eine Filtereinstellung verwendet wird, welche in einer vorhergehenden Sprachpause gewonnen wurde, und/oder das für lange Impulsantworten das NachbildungsfILTER in mehrere Teilfilter aufgespalten wird. Die Erfindung ist insbesondere vorteilhaft für die Befreiung eines Spracheingabesignals in einem Fahrzeug von Störsignalanteilen eines Radios oder dgl., dessen Quellsignal als Referenzsignal zur Verfügung steht.



Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zur Störbefreiung eines Mikrophonsignals.

[0002] Derartige Verfahren gewinnen insbesondere für die Spracheingabe von Kommandos und/oder für Freisprechtelefone zunehmend an Bedeutung, wobei insbesondere die Situation in einem Fahrzeug ein wichtiger Anwendungsfall ist.

[0003] Eine besondere Situation ist in Fahrzeugen häufig dadurch gegeben, daß ein Wiedergabegerät wie z.B. ein Radio, ein Kassetten- oder CD-Abspieler über einen Lautsprecher eine Geräuschumgebung erzeugt, die als Störsignal ein von einem Mikrofon aufgenommenes Sprachsignal, beispielsweise für die Spracherkennung oder Telefonübertragung überlagert. Zur Erkennung von Spracheingaben in einem Spracherkennner oder zur verständlichen Sprachübertragung über Telefon ist das Mikrophonsignal soweit wie möglich von Störsignalanteilen zu befreien.

[0004] Das von einer Störquelle, insbesondere einem Lautsprecher ausgehende Störsignal gelangt nicht nur auf direktem kürzestem Weg zum Mikrofon, sondern tritt auch noch über zahlreiche Reflexionen als eine Überlagerung einer Mehrzahl von Echos mit verschiedenen Laufzeiten im Mikrophonsignal in Erscheinung. Die gesamte Einwirkung des Störsignals von der Störquelle auf das Mikrophonsignal kann durch eine a priori unbekannte Übertragungsfunktion des Raumes, beispielsweise des Fahrgastrumes eines Kraftfahrzeugs beschrieben werden. Die Übertragungsfunktion ändert sich je nach Besetzung des Fahrzeugs und nach Position der einzelnen Personen. Durch Nachbildung dieser Übertragungsfunktion und Filterung eines Referenzsignals von der Störquelle mit dieser Nachbildung kann ein Kompensationssignal erzeugt werden, welches durch Subtraktion vom Mikrophonsignal ein vom Störsignal befreites Signal, beispielsweise ein reines Sprachsignal liefert. Im Realfall stellt die genannte Nachbildung eine mehr oder minder gute Annäherung an die unbekannte Übertragungsfunktion dar und die Störung kann nicht vollständig beseitigt werden.

[0005] Aufgabe der vorliegenden Erfindung ist es, ein Verfahren zur Störbefreiung eines Mikrophonsignals anzugeben, daß bei vertretbarem Signalverarbeitungsaufwand gute Eigenschaften hinsichtlich der Entstörung aufweist.

[0006] Die Erfindung ist im Patentanspruch 1 beschrieben. Die Unteransprüche enthalten vorteilhafte Ausgestaltungen und Weiterbildungen der Erfindung.

[0007] Wesentlich an dem erfundungsgemäßen Verfahren ist, daß die Kompensation des Störsignalanteils im Mikrophonsignal mittels eines aus dem Referenzsignal über die Nachbildung der Übertragungsfunktion erzeugten Kompensationssignals im Frequenzbereich vorgenommen wird, so daß Mikrophonsignal, Kompensationssignal und Ausgangssignal im Frequenzbereich, d.h. in Form von Spektren vorliegen. Die Signalverar-

beitung in diesem Verfahrensschritt im Frequenzbereich erfordert zwar eine spektrale Transformation des Mikrophonsignals, berücksichtigt aber, daß die Nachbildung der Übertragungsfunktion im Frequenzbereich vorteilhafter ist und stellt für eine vorteilhafte nachfolgende zusätzliche Geräuschreduktion des Ausgangssignals, die typischerweise gleichfalls im Frequenzbereich vorgenommen wird, bereits eine besonders geeignete Signalform bereit.

[0008] Durch einfache Näherungen beim Ersatz eines Verarbeitungsschritts mit einem Zeitfenster kann durch Übergang zu einer Faltung im Frequenzbereich eine deutliche Reduzierung des Verarbeitungsaufwands reduziert werden.

[0009] Für lange Impulsantworten der Übertragungsfunktion bzw. deren Nachbildung sieht eine vorteilhafte Weiterbildung der Erfindung eine Aufteilung des Nachbildungsfilters in mehrere Teilverfilter zu zeitversetzten Segmenten des segmentierten Referenzsignals vor, deren Koeffizienten-Aktualisierung zeitlich gestaffelt sein kann, wodurch der Signalverarbeitungsaufwand gering gehalten werden kann.

[0010] Als besonders vorteilhaft erweist es sich, die Entstörung eines Sprachsignals auf der Basis einer Einstellung des Nachbildungsfilters, die in einer vorhergehenden Sprachpause gewonnen und gespeichert wurde, vorzunehmen.

[0011] Die Aufteilung des Nachbildungsfilters in mehrere Teilverfilter und die Störbefreiung auf der Basis einer in einer Sprachpause gewonnenen Filtereinstellung sind auch unabhängig von der Störsignal-Kompensation im Frequenzbereich eigenständig für die Störbefreiung eines Mikrophonsignals realisierbar und vorteilhaft.

[0012] Die Erfindung ist nachfolgend anhand von bevorzugten Ausführungsbeispielen unter Bezugnahme auf die Abbildungen noch eingehend veranschaulicht. Dabei zeigt:

Fig. 1 ein Prinzip der Kompensation eines Radiosignals

Fig. 2a ein Blockschaltbild zu Fig. 1

Fig. 2b ein Blockschaltbild zur Filternachbildung

Fig. 3 ein detailliertes Beispiel zu Fig. 2b

Fig. 4 eine Erweiterung auf mehrere Teilverfilter

Fig. 5 einen Übergang zur Kompensation im Frequenzbereich

Fig. 6 ein detailliertes Beispiel zu Fig. 5b

Fig. 7 ein Ausführungsbeispiel mit mehreren Teilverfiltern

Fig. 8 ein Ausführungsbeispiel mit Speicherung

der Filtereinstellungen

Fig. 9 Signale einer synthetischen Beispielsszene

Fig. 10 Impulsantwort und Übertragungsfunktion zu Fig. 9

Fig. 11 Signal einer ersten Meßszene

Fig. 12 Impulsantwort und Übertragungsfunktion zu Fig. 11

Fig. 13 das Beispiel nach Fig. 11 mit Speicherung der Filtereinstellungen

Fig. 14 eine Sprachpausendetektion zu Fig. 13

Fig. 15 Impulsantworten und Übertragungsfunktionen zu Fig. 11 und Fig. 13

Fig. 16 Übergang von einem Zeitfenster zu einer Faltung im Frequenzbereich

Fig. 17 ein Rechteck-Zeitfenster mit Linienspektrum

Fig. 18 ein Hamming-Zeitfenster mit Linienspektrum

Fig. 19 Staffelung von Signalblöcken bei der Filterberechnung

Fig. 20 Signale einer zweiten Meßszene

Fig. 21 eine Sprachpausendetektion zu Fig. 20

Fig. 22 Impulsantworten und Übertragungsfunktionen zu Fig. 20 und Fig. 21

Fig. 23 Signale einer dritten Meßszene

Fig. 24 eine Sprachpausendetektion zu Fig. 23

Fig. 25 Impulsantworten und Übertragungsfunktionen zu Fig. 23 und Fig. 24

Fig. 26 Signale einer vierten Meßszene

Fig. 27 eine Sprachpausendetektion zu Fig. 26

Fig. 28 Impulsantworten und Übertragungsfunktionen zu Fig. 26 und Fig. 27.

[0013] Fig. 1 stellt das Prinzip einer Einrichtung zur (einkanaligen) Radiosignalcompensation dar. Das vom Lautsprecher abgestrahlte akustische Signal gelangt auf direktem Wege, aber auch über zahlreiche Reflexionen im Fahrzeuginnenraum, auf das Mikrofon des

Spracheingabesystems. Unter der Annahme, daß sich die Übertragungsstrecke G demnach als Transversalfilter mit einer gewichteten Summe zeitlich verzögter Echos darstellt, läßt sich eine Filternachbildung H finden, die im Idealfall $H=G$ eine vollständige Kompensation des Radiosignales ermöglicht.

[0014] Das Lautsprechersignal x wird durch die a priori unbekannte Übertragungsfunktion G des Fahrzeuginnenraumes gefiltert. Es entsteht die Störkomponente r , die sich mit dem Sprachsignal s zu dem Mikrophonsignal y addiert. Um die Störkomponente r zu kompensieren, wird mittels der Filternachbildung H ein Schätzwert r^{\wedge} aus dem Lautsprechersignal x erzeugt. Der Ausgang der Schaltung liefert den Schätzwert für das Sprachsignal:

$$s^{\wedge} = s + r - r^{\wedge} = s+E$$

[0015] Dem Sprachsignal s ist also am Ausgang der Schaltung noch das Fehlersignal $E=r-r^{\wedge}$ überlagert, welches in der Praxis möglichst klein gehalten werden sollte. Das Sprachsignal kann noch Störungen in Form von z.B. Motorgeräuschen oder externen Geräuschen enthalten, die aber in diesem Zusammenhang nicht explizit behandelt werden.

[0016] H ist ein adaptives Filter und arbeitet nach einem in der Literatur bekannten Standardverfahren, dem LMS-Algorithmus (least mean squares). Neben dem Eingangssignal x wird noch das Fehlersignal E benötigt, um die Koeffizientenadaption im Filter H zu bewerkstelligen. Hierfür ist das Ausgangssignal s^{\wedge} der Bestimmung der Filterkoeffizienten zugeführt.

[0017] Fig. 2a zeigt in anderer Darstellung nochmals die Anordnung von Fig. 1 als Radiosignalcompensation. Das adaptive System H kann z.B. im Zeitbereich als FIR-Filter (finite-impulse-response-Filter) realisiert werden. Bei großen Impulsantwortlängen, wie sie in der Praxis häufig auftreten, ist hierzu allerdings ein sehr hoher Rechenaufwand notwendig. Verschiedene Vorteile gegenüber einer Zeitbereichslösung bietet die Realisierung des LMS-Algorithmus im Frequenzbereich (FLMS). Wegen der blockweisen Verarbeitung von Daten in den als diskreten Fouriertransformationen realisierten spektralen Transformationen und der Filterrealisierung im Frequenzbereich durch Multiplikationen wird dieses Verfahren besonders rechenzeitgünstig.

[0018] Fig. 2b zeigt ein Blockschaltbild des FLMS-Algorithmus. Die zugehörige Theorie ist an sich bekannt und daher an dieser Stelle nicht im Detail behandelt. Es bedeuten F eine spektrale Transformation FFT eines Zeitsignals in den Frequenzbereich und F^{-1} die inverse IFFT. Die als Projektionen P_1 , P_2 und P_3 bezeichneten Verarbeitungsschritte dienen der korrekten Segmentierung der Daten durch die blockweise Verwendung mit der FFT bzw. IFFT und werden später noch genauer erläutert. Die Arbeitsweise des Filters besteht in der Multiplikation des Referenzspektrums X mit dem Filter-Koeffizientenvektor H . Das Spektrum des

Filterausgangs R^A wird über F^{-1} zurück in den Zeitbereich transformiert. Nach Anwendung der Projektion P2 auf den Realteil des so erhaltenen Kompensationssignals steht das Signal r^A zur Verfügung. Die Differenz der Signale

$$s^A = y - r^A = s + r - r^A = s + E$$

stellt den eigentlichen Ausgang, eine Schätzung der Spracheingabe, dar.

[0019] Wesentlicher Bestandteil des adaptiven Filters ist die Koeffizientenadaption im Block K, die im Fig. 2b durch die Erneuerungsgleichung

$$H' = H' + \Delta H'$$

beschrieben wird. Die hier mit zwei spektralen Transformationen besonders aufwendige Projektion P1 berechnet aus H' den für die Filterung benötigten Koeffizientenvektor H . Zur Berechnung des Korrekturvektors $\Delta H'$ wird neben dem Referenzspektrum X das Spektrum S^A des mit P_3 bewerteten Ausgangssignales $s + r - r^A$ benötigt.

[0020] Ein detailliertes Blockschaltbild des in Fig. 2b dargestellten FLMS-Algorithmus zeigt Fig. 3. Die Abtastwerte eines Signals und die Stützstellen der FFT seien in gebräuchlicher Weise als samples bezeichnet. Alle Spektraltransformationen und deren Inverse sind als 256-Punkte - FFT's, die jeweils um 128 samples überlappen, zu segmentieren. Zu beachten ist, daß sich das Ausgangssignal s^A im Zeitbereich aus 128-sample-Blöcken zusammensetzt. Es entsteht aus der Differenz der zweiten Blockhälften (also jeweils der samples 129 bis 256) von Mikrofonsignal und gefiltertem Kompensationssignal r^A . Aufwendig ist die Projektion P1, die 2 FFT's benötigt und den Vektor H' in den Vektor H umrechnet. Hierbei wird aus dem komplexen 256-Punkte-Ergebnisvektor der Rücktransformation vom Frequenz- in den Zeitbereich (IFFT) die erste Hälfte (samples 1 bis 128) ausgeschnitten und die zweite Hälfte (samples 129 bis 256) zu Null gesetzt. Nach Anwendung dieses Rechteckfensters im Zeitbereich erfolgt wieder mittels FFT die Transformation in den Frequenzbereich. Einfach ist die Projektion P2. Sie besteht aus der oben schon beschriebenen Ausschnittsbildung der letzten 128 samples, wodurch aus überlappenden 256-sample-Blöcken wieder nicht überlappende 128-sample-Blöcke entstehen. Ebenfalls sehr einfach ist schließlich auch die Projektion P3, welche umgekehrt aus nicht überlappenden 128-sample-Blöcken des Ausgangssignals durch Voranstellen von 128 Nullwerten wiederum überlappende 256-sample-Blöcke bereitstellt. Die Adaption der Filterkoeffizienten H'_{L+1} für einen Zyklus $L+1$ besteht aus der Addition eines Erneuerungsvektors $\Delta H'_L$ zum alten Koeffizientenvektor H'_L . Diese Erneuerung errechnet sich aus dem Produkt zwischen dem Spektrum S^A_L des Ausgangssignals und dem konjugiert komplexen Spektrum X^*_L des Referenz-

signales - gewichtet mit einer spektralen Leistungsnormierung $2\mu_L$, $\Delta H'_L = 2\mu_L \cdot X^*_L \cdot S^A_L$. Zum Zweck dieser Leistungsnormierung ist der mit einer Konstanten 2α multiplizierte Kehrwert des geglätteten Referenzleistungsspektrums $S_{xx,L}$ zu berechnen $2\mu_L = 2\alpha / S_{xx,L}$, wozu ein rekursives Filter 1. Ordnung mit einer Konstanten β dient

$$S_{xx,L} = \beta \cdot |X_L|^2 + (1-\beta) \cdot S_{xx,L-1}$$

[0021] Die Arbeitsweise des LMS-Algorithmus wird erheblich von der Adaptionskonstante α und der Glättungskonstante β beeinflußt. Zwischenspeicher in Rekursionsschleifen sind mit Sp bezeichnet.

[0022] Die bisher beschriebene Anordnung des FLMS-Algorithmus erlaubt Filternachbildungen mit einer maximalen Impulsantwortlänge von einer halben FFT-Länge, im Beispieldfall also 128 samples. Sollen längere Impulsantworten kompensiert werden, ist der schon bekannte FLMS-Algorithmus für einen Teilfilter (Fig. 4a) auf n Teilfilter zu erweitern. Eine 3-Teilfilter-Lösung mit einer Impulsantwortlänge von $3 \cdot 128 = 384$ samples hat sich bei der Radiosignalunterdrückung im Pkw mit einem Spracheingabesystem bewährt (Fig. 4b). Der im Fig. 4a mit B bezeichnete Block mit den Eingangssignalen X und S^A und dem Kompensations-Spektrum R^A als Ausgang ist durch die im Fig. 4b dargestellte Erweiterung zu ersetzen. Das Spektrum X des Referenzsignals wird durch Zwischenspeicher D um 1 bzw. 2 Blocklängen verzögert und das unverzögerte X_1 und die beiden verzögerten Spektren X_2, X_3 werden separat in mit in einer erweiterten Projektion P1 getrennt bestimmten Koeffizientenvektoren H_1, H_2, H_3 multipliziert. Die Bildung der Koeffizientenvektoren erfolgt analog zum Fall nur eines Teilfilters, wobei in K_1, K_2, K_3 jeweils das zugehörige Referenzspektrum mit dem Spektrum S^A des Ausgangssignals verknüpft wird. Der Aufwand wird hauptsächlich durch die Verdreifachung der Projektion P1 beträchtlich erhöht. Zusätzlicher Speicherplatzbedarf wird notwendig um die Spektren des um 1 bzw. 2 Blocklängen zeitlich älteren Referenzsignals X bereitzustellen.

[0023] Bei der beispielhaft angegebenen Aufgabenstellung der Unterdrückung des Radiosignals bei Spracheingabe im Kfz ist es vorteilhaft die Ausgangsdaten nicht im Zeit- sondern im Frequenzbereich auszugeben, da dadurch eine verbesserte Anpassung an eine nachgeschaltete Geräuschunterdrückung erreicht werden kann. Der bereits vorgestellte FLMS-Algorithmus mit einem Teilfilter benötigt gemäß Fig. 5a insgesamt 5 FFT's bei einem Ausgangssignal im Zeitbereich. Wird dem Ausgang eine FFT nachgeschaltet, erhöht sich der Aufwand bei einem Frequenzbereichs-Ausgangssignal auf 6 FFT's. Die gleiche FFT-Anzahl ergibt sich zunächst auch bei einer äquivalenten Lösung nach Fig. 5b. Diese Variante besitzt jedoch folgende Vorteile:

- Bei der zeitgleichen Spektralanalyse der Signale x

und y ist nur eine einzige 256-Punkte-FFT mit geringem Zusatzaufwand für eine spektrale Separation notwendig. Man erzielt eine Einsparung von 1 FFT.

- Die hier mit P4 gekennzeichnete und neu definierte Projektion ist bis auf das verwendete Zeitfenster formal identisch mit der Projektion P1. Wie später gezeigt wird, läßt sich P4 durch eine relativ einfache Faltungsoperation im Frequenzbereich ersetzen, ohne daß eine merkliche Einbuße an Qualität in Kauf genommen werden muß. Man erzielt eine Einsparung von 2 FFT's.

[0024] Fig. 6 stellt ein ausführlicheres Blockschaltbild des FLMS-Algorithmus mit Frequenzbereichs-Ausgangssignal dar und gestattet nochmals einen Vergleich mit Fig. 3 (Zeitbereichs-Ausgang). Unverändert geblieben ist die Filteradaption bestehend aus Glättung der spektralen Leistung, Leistungsnormalierung und Koeffizientenerneuerung. Neu sind die FFT im Mikrofonkanal, die Differenzbildung $Y-R^A$ im Frequenz- statt im Zeitbereich zur Ausgangsbildung, und schließlich die neu definierte Projektion P4, welche sich lediglich durch das komplementäre Zeitbereichsfenster von der Projektion P1 unterscheidet.

[0025] Als Vorstufe einer nachfolgend beschriebenen bevorzugten Ausführung betrachte man Fig. 7. Dargestellt ist der FLMS-Algorithmus mit 3 Teilstufen (384-sample-Impulsantwort), welcher eine ausreichende Unterdrückung des Radiosignales im Mikrofonkanal des Spracheingabesystems liefert. Die Projektionen P1 und P4 sind vereinfacht dargestellt. Es ist der schon aus Fig. 4b bekannte zusätzliche Aufwand in Form der Speicher P sowie die Verdreifachung der Projektion P1 ersichtlich. Im Gegensatz zur 1-Teilstufen-Lösung nach Fig. 6 wird die Summe W von gegenwärtigem und den beiden zeitlich vorausgegangenen Referenz-Leistungsspektren auf den Eingang des rekursiven Filters gegeben. Die Tatsache, daß am Filterausgang nun praktisch die 3-fache geglättete spektrale Leistung vorliegt, wird nach der Kehrwertbildung durch Multiplikation mit der Konstanten 6α berücksichtigt. Nach der spektralen Leistungsnormalierung des in P4 modifizierten Ausgangsspektrums S^A wird die Filter-Adaption nun für die 3 Koeffizientenvektoren der 3 Teilstufen separat durchgeführt.

[0026] Ein Beispiel Z0 für die Wirkungsweise der Erfindung nach Fig. 7 zeigt Fig. 9. Die Eingangsdaten wurden synthetisch generiert. Das Referenzsignal X stellt 100000 Abtastwerte eines weißen Gaußrauschen bei einer Abtastfolgefrequenz von $fs=12$ kHz dar. Das Mikrophonsignal Y entstand durch Faltung dieses Rauschsignals mit einer ebenfalls konstruierten 384-sample-Impulsantwort sowie der Addition eines äußerst schwachen Sprachsignals. Beim Abhören dieses in Fig. 9 oben aufgezeichneten Signals y sind die 10 gesprochenen Ziffern gerade noch im farbigen (weil

gefilterten) Rauschen zu erkennen. Das in den Zeitbereich zurücktransformierte Ausgangssignal des Schätzers befreit nach einem ca. 1 Sekunde (12000 samples) dauernden Einschwingvorgang sehr wirkungsvoll die Spracheingabe vom Rauschen und liefert ein unverzerrtes jedoch leicht verhalltes Sprachsignal S^A (Fig. 9 unten). Die verwendeten beiden Parameter waren $\alpha=0,05$ und $\beta=0,5$, Werte, die sich auch bei den später vorgestellten Beispielen gut bewährt haben.

[0027] Aus den jeweils 129 samples langen Teilkoeffizientenvektoren H_1, H_2, H_3 der 3 Teilstufen nach Bild 7 läßt sich nun zu jedem Zeitpunkt die hieraus resultierende 3*128-sample-Impulsantwort bzw. die zugehörige Filterübertragungsfunktion berechnen. So zeigt Fig. 10 oben die 384-sample-Impulsantwort, wie sie sich ganz am Ende der Szene - also nachdem die Ziffer "0" gesprochen wurde - ergibt. Sie ist ein sehr genaues Abbild derjenigen Impulsantwort, die zur Faltung mit weißem Gaußrauschen und damit zur synthetischen Generierung des Signales mikro benutzt wurde. Die zugehörige Betragsübertragungsfunktion (Fig. 10 unten) im Bereich zwischen den Frequenzen 0 und $fs/2 = 6$ kHz stellt einen mit zahlreichen schmalbandigen Resonanzüberhöhungen behafteten Tiefpaßfrequenzgang dar.

[0028] Weißes Rauschen als Referenzeingangssignal und gefiltertes "farbiges" Rauschen als Mikrofoneingangssignal stellen im Sinne der Aufgabenstellung, eine Nachbildung dieses Filters zu finden, den einfachsten Fall dar. Da das Referenzsignal per Definition alle Frequenzanteile enthält, gelingt die Filteradaption hier am schnellsten. Die zusätzliche additive Spracheingabe im Mikrofoneingangssignal - also das eigentliche Nutzsignal des Spracheingabesystems - stellt für den (F)LMS-Algorithmus eine Störung dar, welche die korrekte Adaption der Filterkoeffizienten behindert. Anders ausgedrückt: nur in Sprachpausen ist das System dazu in der Lage die Raumakustik des Fahrzeuginnenraums (Strecke Radiolautsprecher zum Mikrofon) richtig nachzubilden und dadurch eine Kompensation der Radiowiedergabe zu bewirken. Im oben demonstrierten Beispiel gemäß Fig. 9 gelingt dies sehr gut, da der Mikrofon-eingang im wesentlichen aus Rauschen und nur zu einem sehr geringen Teil aus Spracheingabe besteht.

[0029] Aus echten Messungen im Fahrzeug hingegen entstammten das an den Radiolautsprecherklemmen abgegriffene Referenzsignal radio und das vom Mikrofon des Spracheingabesystems aufgezeichnete Signal mikro der Szene Z1. Dieses Mikrophonsignal ist in Fig. 11 oben dargestellt, besteht aus 100000 samples und besitzt demnach bei einer Sampling-Frequenz von 12 kHz eine zeitliche Dauer von ca. 8,3 Sekunden. Es handelt sich um flüssig und relativ schnelle gesprochene Sprache eines hinten rechts im Auto sitzenden Fahrzeuginsassen, während gleichzeitig Musik mit normaler Lautstärke aus dem Autoradio-Lautsprecher erklingt. Nach Anwendung der Entstörmaßnahme gemäß Fig. 7

und Umrechnung in den Zeitbereich ergibt sich das in Fig. 11 unten dargestellte Ausgangssignal. Der Hörtest ergibt eine deutliche Herausarbeitung des Sprachanteils bzw. eine vor allem in den kurzen Sprachpausen bemerkenswerte Musikunterdrückung. Auffällig und von Nachteil ist jedoch, daß die erwünschte Radiosignalunterdrückung in starkem Maße davon abhängt, ob gerade gesprochen wird oder nicht. Die wieder am Szenenende ermittelte 384-sample-Impulsantwort mit zugehöriger Übertragungsfunktion ist aus Fig. 12 ersichtlich. Eine korrekte Impulsantwort ist an den typischen Nullsamples (Totzeit) am Anfang zu erkennen, welche von der Laufzeit des Direktschalls vom Radio-lautsprecher zum Mikrophon herrühren. Aus den hier vorhandenen starken Störungen am Anfang sowie am Ende der Impulsantwort läßt sich demnach der Schluß ziehen, daß die Filteradaption an dieser Stelle wegen vorhandener Spracheingabe äußerst unzureichend ist.

[0030] Die im folgenden anhand von Fig. 8 beschriebene Ausführungsform beruht auf folgender Grundidee: ein geeignetes Merkmal dient zusammen mit einem Schwellenwert als Indikator für eine Spracheingabe. Unterschreitet das Merkmal die Schwelle, so ist dies ein Anzeichen für fehlende Spracheingabe. In diesem Fall kann - wie oben schon festgestellt - eine weitgehend ungestörte Filteradaption erfolgen. Bei Spracheingabe wird nun auf denjenigen Filterkoeffizientensatz zurückgegriffen, der unmittelbar vor der Schwellenüberschreitung - d.h. am Ende der vorangegangenen Sprachpause - abgespeichert wurde. Diese gespeicherten Koeffizienten H_{10}, H_{20}, H_{30} liefern im Regelfall eine deutlich bessere Radiosignal-Kompensation als die unter dem störenden Einfluß der Spracheingabe sich ständig ändernden aktuellen Koeffizienten H, H_2, H_3 .

[0031] Fig. 8 stellt eine Ausführung mit einer weiter verbesserten FLMS-Verarbeitung mit 3 Teilstufen dar. Neben den schon in Fig. 7 vorhandenen aktuellen Filterkoeffizientenvektoren H_1, H_2, H_3 , welche zur Bildung des fortlaufend adoptierten Ausgangssignals $y-R$ benötigt wurden, existiert nun ein zusätzliches Ausgangssignal ($y-Ro$), das unter Verwendung gespeicherter Koeffizienten H_{10}, H_{20}, H_{30} gebildet ist. Die aktuellen Koeffizientensätze H_1, H_2, H_3 stellen nur bei fehlender Spracheingabe im eingeschwungenen Zustand ein brauchbares Kompensationsfilter im Frequenzbereich dar, liefern hingegen bei Spracheingabe ungenügende Filtereigenschaften, weil der Adoptionsprozeß in der Regelschleife ständig gestört wird. Bei fehlender Spracheingabe d.h. hoher Filterqualität sind die drei Schalter geschlossen und es werden die aktuellen Koeffizientensätze in die Koeffizientenspeicher M_1, M_2, M_3 geschrieben: $H_{10}=H_1, H_{20}=H_2, H_{30}=H_3$. Die Ausgänge ($y-Ro$) und ($y-Ra$) sind identisch. Einsetzende Spracheingaben bewirken ein Öffnen der 3 Schalter, wodurch die zuletzt in den Speichern M_1, M_2, M_3 befindlichen Koeffizienten H_{10}, H_{20}, H_{30} nicht mehr überschrieben werden und unverändert bleiben. Dieser

Zustand, in welchem sich die Ausgänge ($y-Ro$) und ($y-Ra$) unterscheiden, wird solange beibehalten, bis wieder eine Sprachpause detektiert und die Schalter geschlossen werden.

- 5 **[0032]** Als Sprachpausenmerkmal fea hat sich die geglättete Summe aller Absolutwerte der Koeffizientenkorrekturvektoren $\Delta H_1', \Delta H_2', \Delta H_3'$ bewährt (Fig. 8a). Diese Größe ist gleich Null bzw. weist kleine Zahlenwerte auf, wenn es keinen oder nur einen geringen Bedarf gibt, die Koeffizienten abzuändern. In Sprachpausen ist dies der Fall, der Regelkreis ist praktisch eingeschwungen. Störungen, wie sie durch Spracheingabe - aber auch durch Bewegungen der Fahrzeuginsassen - hervorgerufen werden, haben einen erhöhten Nachregelbedarf zur Folge, was sich durch entsprechend große Zahlenwerte bei $\Delta H_1', \Delta H_2', \Delta H_3'$ und somit beim Merkmal fea bemerkbar macht. Ein Glättungsfilter beispielsweise ein rekursiver Tiefpaß 1. Ordnung mit dem Eingang $feat$ stellt an seinem Ausgang das geglättete Sprachpausen-Merkmal fea zur Verfügung, welches nach Vergleich mit einem Schwellenwert th die Schalter für die Koeffizientenübernahme steuert.
- 10 **[0033]** Die Wirkungsweise des verbesserten FLMS-Algorithmus nach Fig. 8 demonstriert Fig. 13. Oben ist das aufgezeichnete Signal y der Szene Z1 (vgl. Fig. 11 oben) dargestellt, unten das gewonnene Ausgangssignal. Schon der visuelle Vergleich der Ausgangssignale von Fig. 13 und Fig. 11 zeigt die verbesserte Herausarbeitung der Sprachpassagen. Der vergleichende Hörtest bestätigt dies: auch während der Spracheingabe ist die Musikunterdrückung deutlich besser. Den Verlauf des Sprachpausenmerkmals und der konstanten Schwelle über der Zeit (hier in FFT-Blöcken skaliert) zeigt Fig. 14 oben. In den durch die Schwellenunterschreitung detektierten Sprachpausen (Fig. 14 unten) findet laufend die Übernahme der Koeffizienten in die Speicher wie beschrieben statt, um dort während der Spracheingabe als gespeicherte Koeffizienten zur Verfügung zu stehen. Die schon in Fig. 12 am Szenenende gemessene 384-sample-Impulsantwort mit zugehöriger Betragsübertragungsfunktion ist in Fig. 15 als aktuelle Impulsantwort (a) bzw. aktuelle Übertragungsfunktion (b) dargestellt. Im Gegensatz zu dieser infolge Spracheingabe stark gestörten Schätzung aus den aktuellen Koeffizienten H_1, H_2, H_3 ist aus den gespeicherten Koeffizienten H_{10}, H_{20}, H_{30} eine Impulsantwort (c) und eine Übertragungsfunktion (d) hoher Qualität berechenbar. Die Impulsantwort aus den gespeicherten Koeffizienten weist die typischen Nullsamples am Anfang auf, welche durch die Laufzeit des Direktschalls vom Radio-lautsprecher zum Spracheingabemikrophon verursacht werden. Aus der im Beispieldfall abzulesenden Totzeit von ca. 40 samples läßt sich die Entfernung zwischen Lautsprecher und Mikrophon bestimmen.
- 15 **[0034]** Wie vorstehend schon angedeutet läßt sich die aufwendige Projektion P4 (IFFT, Fenster rechts im Zeitbereich, FFT) ohne merkliche Einbuße an Qualität

durch eine relativ einfache Faltung im Frequenzbereich ersetzen, wodurch 2 FFT's eingespart werden. Dazu betrachte man Fig. 16. In einem ersten Schritt wird das "rechtsseitige" 128-sample-Rechteckfenster im Zeitbereich (Fig. 16a) bei der idealen Projektion ersetzt durch ein 128-sample-Hammingfenster (Fig. 16b). Gegenüber dem Rechteckfenster besitzt dieses den Vorteil eines bedeutend schmäleren Spektrums. Wie Fig. 17 zeigt, besteht beim Rechteckfenster der Realteil des Spektrums aus einer einzigen Linie (Gleichanteil), während das zur Mitte antisymmetrische Imaginärteil-Spektrum aus vielen nach außen hin langsam abfallenden Linien mit alternierenden Nullstellen besteht. Im Gegensatz dazu beschränkt sich das komplexe Spektrum des Hammingfensters (Fig. 18) auf insgesamt 7 Linien, von denen im symmetrischen Realteil nur 3 und im antisymmetrischen Imaginärteil nur 4 Werte von Null verschieden sind. Sämtliche weiter außen liegenden Anteile sind vernachlässigbar gering. Diese spezielle Eigenschaft des Hammingfensters ermöglicht es vorteilhafterweise die Multiplikation im Zeitbereich (Fig. 16b) zu ersetzen durch eine Faltung mit dem zugehörigen 7-sample-Spektrum im Frequenzbereich und damit eine IFFT und eine FFT einzusparen (Fig. 16c).

[0035] Prinzipiell läßt sich natürlich auch die Projektion P1 (IFFT - linksseitiges Rechteckfenster - FFT) ersetzen durch eine entsprechende Faltungsoperation im Frequenzbereich mit dem konjugiert komplexen 7-Linien-Spektrum. Experimente haben jedoch gezeigt, daß Einsparungen an dieser Stelle erkauft werden mit einer deutlichen Verschlechterung des Einschwingverhaltens. Aufwandsgunstige Lösungen lassen sich trotzdem dadurch erzielen, daß in dem LMS-Algorithmus nach Fig. 8 die 3 Projektionen P1 nicht gleichzeitig in einem 256-sample-Inputdatenblock abgearbeitet werden müssen. Die mit 128-samples überlappenden Inputdatenblöcke der Länge 256 sind mit einer willkürlich bei "1" beginnenden Nummerierung in Fig. 19a skizziert. So ist es z.B. möglich bei modulo-3-Zählweise der Inputdatenblöcke die 3 Teilfilterprojektionen nicht parallel (Fig. 19b) sondern sequentiell in aufeinanderfolgenden Blöcken Fig. 19 zu berechnen. Dadurch sind bei idealer Projektion P1 pro Datenblock nicht 6 sondern nur noch 2 FFT's notwendig. Es hat sich gezeigt, daß die Kompensation des Radiosignales auch noch ausreichend funktioniert, wenn die Abstände zwischen den zu berechnenden Teilfilter-Projektionen noch größer gewählt werden. Zählt man die Blöcke z.B. modulo 6, so ist lediglich in jedem zweiten Block eine Projektion zu berechnen (Fig. 19d). Selbst eine Reduzierung auf einen Abstand von vier Blöcken zwischen zwei aufeinanderfolgenden P1 Berechnungen mittels modulo-12-Zählung führt noch zu brauchbaren Ergebnissen (Fig. 19e).

[0036] Die Leistungsfähigkeit des FLMS-Algorithmus mit 3 Teilfiltern gemäß Blockschaltung Fig. 8 und einer sequentiellen Berechnung der idealen Projektion P1 im Zeitraster nach Fig. 19e sowie der Projektion P2 mittels

Faltung im Frequenzbereich (Fig. 16c) mit einem komplexen 7-Linien-Spektrum (Fig. 18) sei anhand von 3 Meßszenen demonstriert.

[0037] Die erste dieser Szenen Z2 beinhaltet Spracheingabe von Ziffern, wobei der Radiolautsprecher annähernd weißes Rauschen mit verhältnismäßig hoher Lautstärke abstrahlt. Das zugehörige 100000-sample-Mikrophonsignal ist in Fig. 20 oben, das extrahierte Ausgangssignal in Fig. 20 unten dargestellt. Eine deutliche Rauschbefreiung des Outputsignals gegenüber dem Mikrophoninput stellt man durch Abhörvergleich fest. Der zeitliche Verlauf des Sprachpausenmerkmals ist zusammen mit der konstanten Schwelle th Fig. 21 oben abgebildet und die hieraus abgeleiteten Sprachpausen bzw. die zugeordneten Schalterstellungen in Fig. 21 unten. Schließlich zeigt Fig. 22 in zu Fig. 15 analoger Weise die am Szenenende gefundene Impulsantwort (a) und Übertragungsfunktion (b) auf der Basis der aktuellen Koeffizienten und die entsprechenden Größen (c), (d) auf der Basis der Sprachpauseneinstellung. Es ist deutlich erkennbar, daß die am Szenenende gefundene aktuelle Impulsantwort ein infolge Spracheingabe gestörtes Ergebnis darstellt, während die aus der letzten Sprachpause stammende Impulsantwort aus den gespeicherten Koeffizientensätzen eine hohe Qualität aufweist.

[0038] Die ersten 100000 samples einer Meßszene Z3 mit POP-Musik im Radio und flüssig bis schnell gesprochener Sprache der rechts hinten sitzenden Person sind in Form des Mikrophonsignales y in Fig. 23 oben aufgezeichnet. Nach ca. 10000 samples (0,83 s) wird das Radiosignal brauchbar unterdrückt (Fig. 23 unten). Auch bei der im letzten Drittel dieser Szene einsetzenden Spracheingabe bleibt die POP-Musikunterdrückung wirksam erhalten, wodurch die Sprachverständlichkeit hier gegenüber dem Mikrophonsignal merklich verbessert wird. Nach einer langen Sprachpause kommt es wegen der anschließenden pausenfreien Spracheingabe nicht mehr zu einer Schwellenunterschreitung (Fig. 24). Aus diesem Grunde ist die in Fig. 25 unten am Ende der Szene festgehaltene Impulsantwort auf der Basis der gespeicherten Koeffizienten zeitlich relativ veraltet, weil sie bereits ca. 2,3 s vorher aktuell war (215 Blöcke * 10,7 ms). Wieder weist die aktuelle Impulsantwort (Fig. 25 oben) starke von der Spracheingabe herrührende Störungen auf. Wie ein Vergleich mit der ähnlichen Szene Z1 nach Figuren 11 bis 15 zeigt, ist trotz des stark verringerten Rechenaufwandes die Qualität der Störbefreiung unverändert hoch.

[0039] Die letzte Szene Z4 nach Fig. 26 wurde ohne Spracheingabe erstellt und soll abschließend nochmals die Musikunterdrückungseigenschaften des beschriebenen FLMS-Algorithmus demonstrieren. Nach ca. 18000 samples bzw. 1,5 s wird - wie aus Fig. 26 unten ersichtlich - die Musik wirksam unterdrückt. Diese Eigenschaft wird bis zum Szenenende mit unveränderter Qualität beibehalten. Fig. 27 zeigt auf, daß das

Sprachpausen-Größe fea überwiegend unter der Schwelle th bleibt. Die Zeiten, in welchen auf die gespeicherten Koeffizienten zurückgegriffen wird, sind demnach nur sehr kurz. Impulsantwort und Übertragungsfunktion aus aktuellen Koeffizienten sind daher im wesentlichen mit den entsprechenden Verläufen aus Sprachpausen-Koeffizienten identisch.

Patentansprüche

1. Verfahren zur Störbefreiung eines Mikrophonsignals von Anteilen eines Quellsignals, daß als Referenzsignal (x) vorliegt und nach Durchlaufen einer Übertragungsstrecke mit a priori unbekannter Übertragungsfunktion (G) sich im Mikrophonsignal als Störsignal (r) einem Sprachsignal (s) überlagert, durch adaptive Nachbildung des Störsignals und Kompensation des tatsächlichen und des nachgebildeten Störsignals in einem Ausgangssignal, wobei das Mikrophonsignal gleichfalls in den Frequenzbereich transformiert, die Kompensation an Signalen im Frequenzbereich vorgenommen und das im Frequenzbereich vorliegende Ausgangssignal zur Adaption der Nachbildung mit dem im Frequenzbereich vorliegenden Referenzsignal verknüpft wird.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das Ausgangssignalspektrum in den Zeitbereich transformiert, das Zeitsignal durch Voranstellen von Nullwerten auf doppelte Länge gebracht, in den Frequenzbereich rücktransformiert und der Nachbildung der Übertragungsfunktion zugrundegelegt wird.

3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das Ausgangssignalspektrum mit dem Spektrum eines Hamming-Zeitfensters gefaltet wird und der Nachbildung der Übertragungsfunktion zugrundegelegt wird.

4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß zur Nachbildung des Störsignalanteils eine adaptive Filterfunktion eines Nachbildungsfilters auf das Referenzsignal angewandt wird.

5. Verfahren nach Anspruch 4, dadurch gekennzeichnet, daß die Filterfunktion durch einen Koeffizientenvektor vorgegeben wird, dessen Koeffizienten adaptiv eingestellt werden.

6. Verfahren nach Anspruch 4 oder 5, dadurch gekennzeichnet, daß das Auftreten eines Sprachsignals im Mikrophonsignal detektiert wird und bei Auftreten eines Sprachsignals die vor Auftreten des Sprachsignals eingestellte Filterfunktion zur Bildung des Ausgangssignals beibehalten wird.

7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, daß auch bei Detektion eines Sprachsignals die adaptive Nachführung einer aktuellen Filterfunktion zusätzlich zur Bildung des Ausgangssignals fortgeführt wird.

8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß das Auftreten eines Sprachsignals aus einer Veränderung der aktuellen Filterfunktion detektiert wird.

9. Verfahren nach Anspruch 8, dadurch gekennzeichnet, daß die Veränderung der aktuellen Filterfunktion für die Detektion des Auftretens eines Sprachsignals zeitlich geglättet wird.

10. Verfahren nach einem der Ansprüche 4 bis 9, dadurch gekennzeichnet, daß die Filterfunktion in mehrere Teilverfilterfunktion zu aufeinanderfolgenden Abschnitten einer Gesamt-Impulsantwort aller Teilfilter aufgespalten ist und auf Referenzsignalspektren zu zeitlich versetzten Zeitsegmenten des segmentierten Referenz-Zeitsignals angewandt wird.

11. Verfahren nach Anspruch 10, dadurch gekennzeichnet, daß die Adaption der Filterfunktion für die Teilverfilter parallel durchgeführt wird.

12. Verfahren nach Anspruch 10, dadurch gekennzeichnet, daß die Adaption der Filterfunktion für die einzelnen Teilverfilter zeitsequentiell durchgeführt wird.

35

40

45

50

55

55

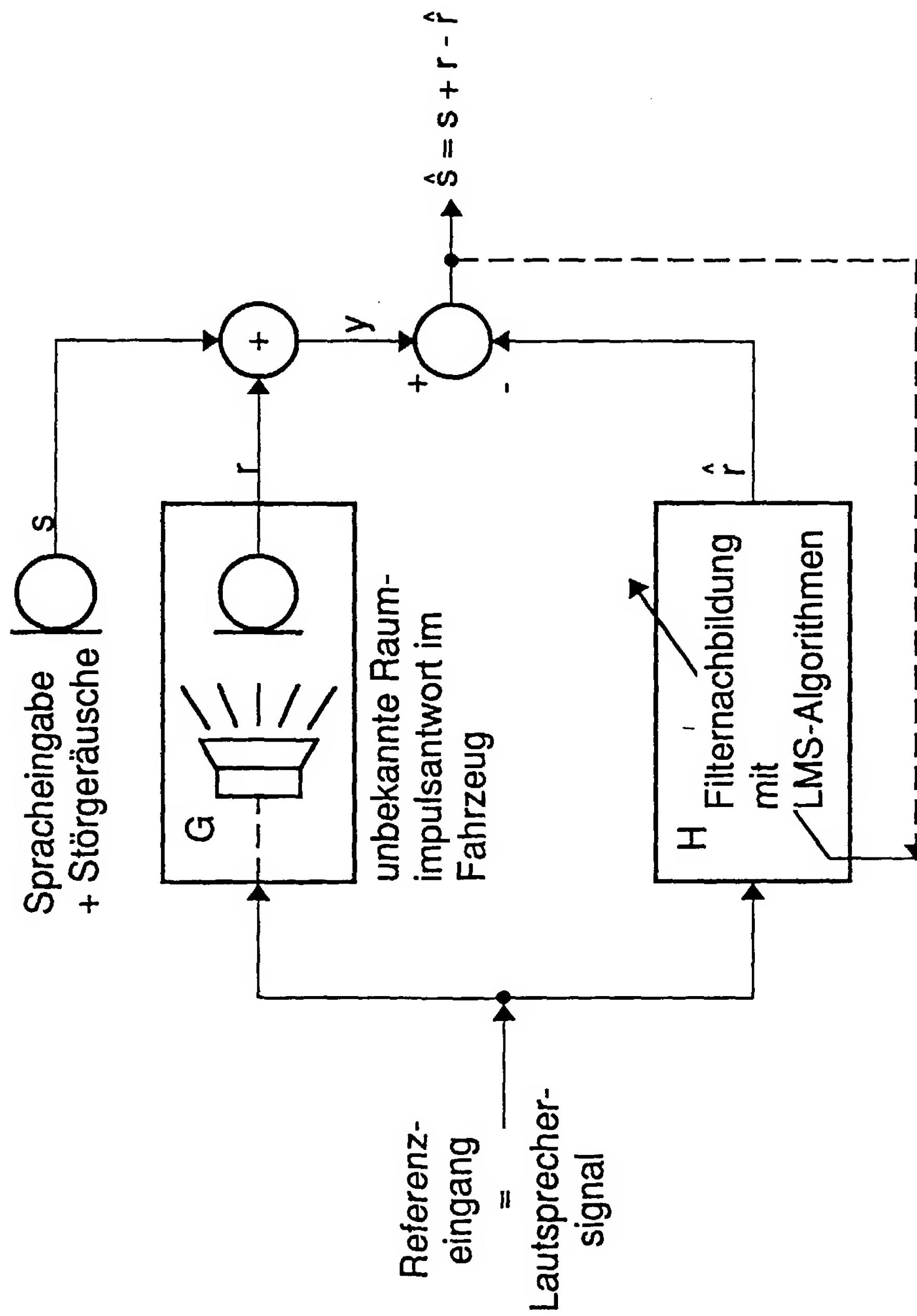


FIG. 1

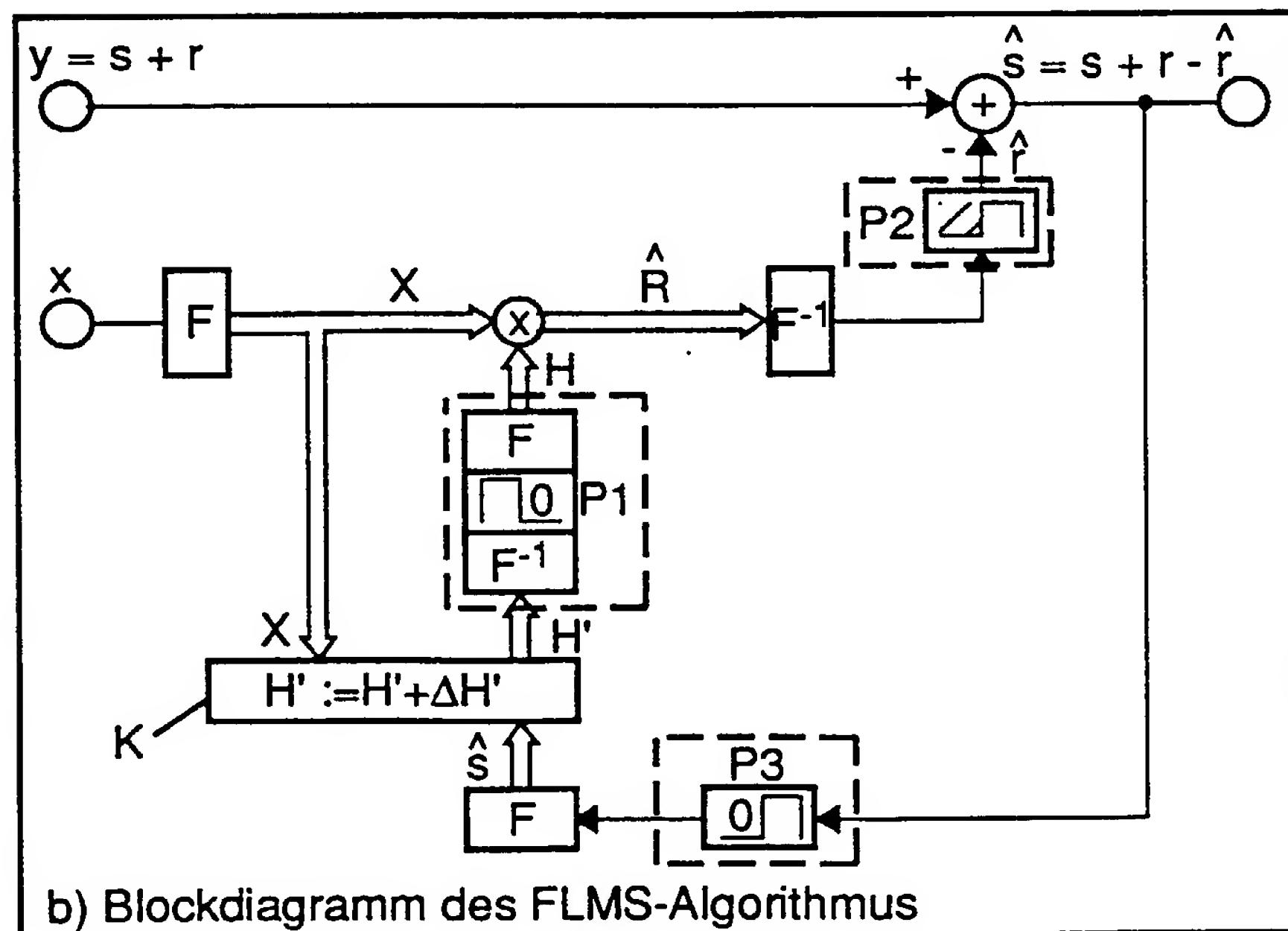
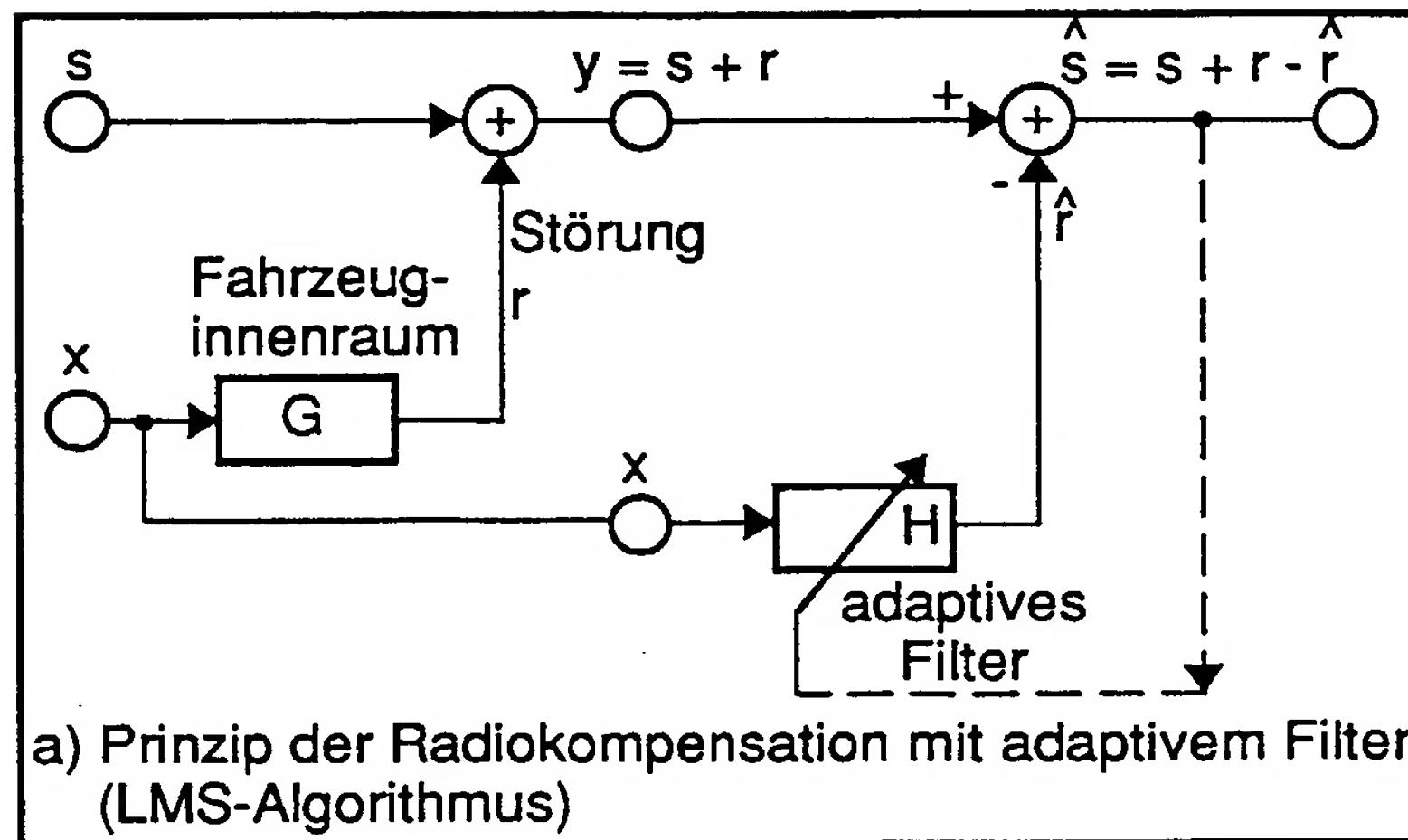


FIG.2

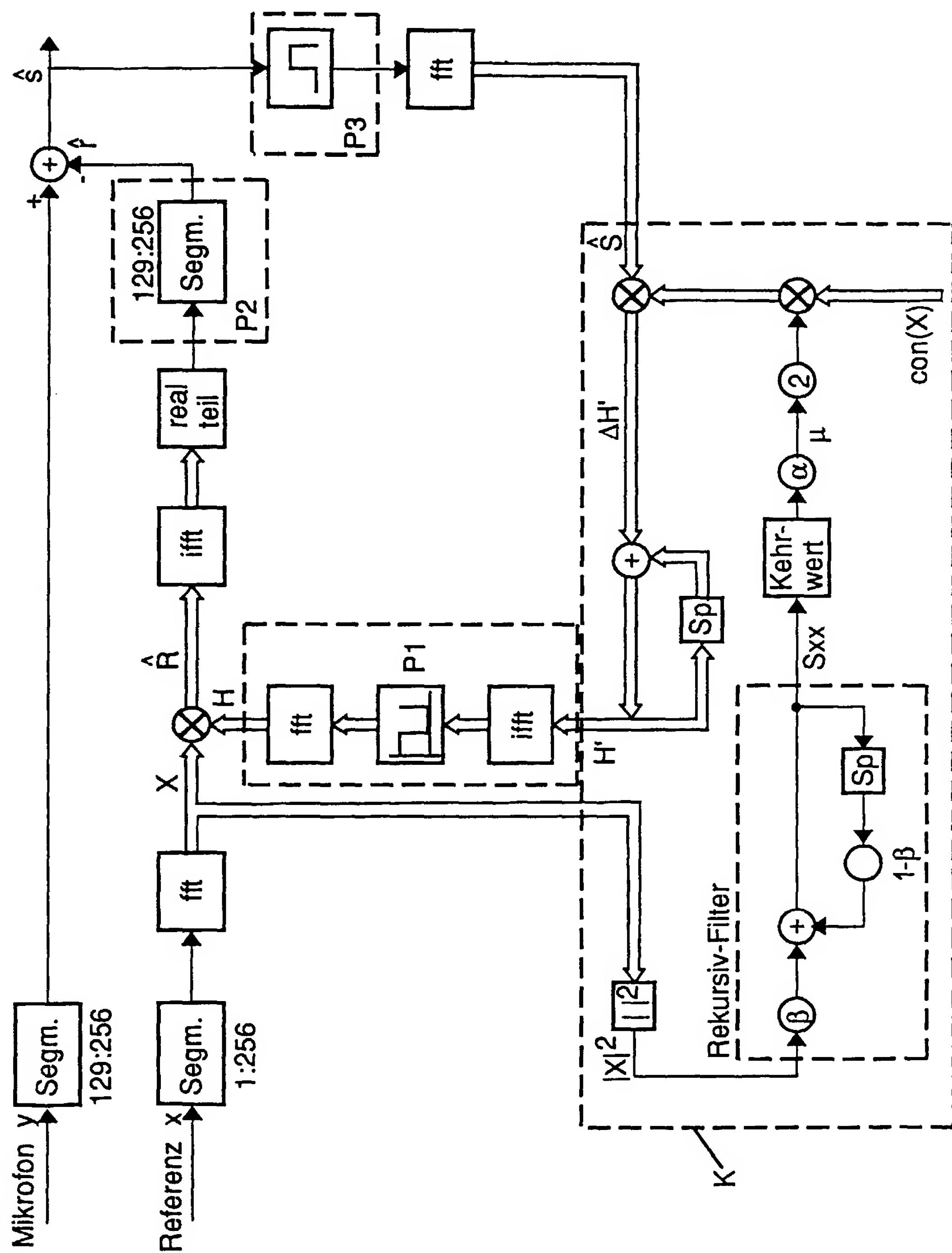


FIG.3

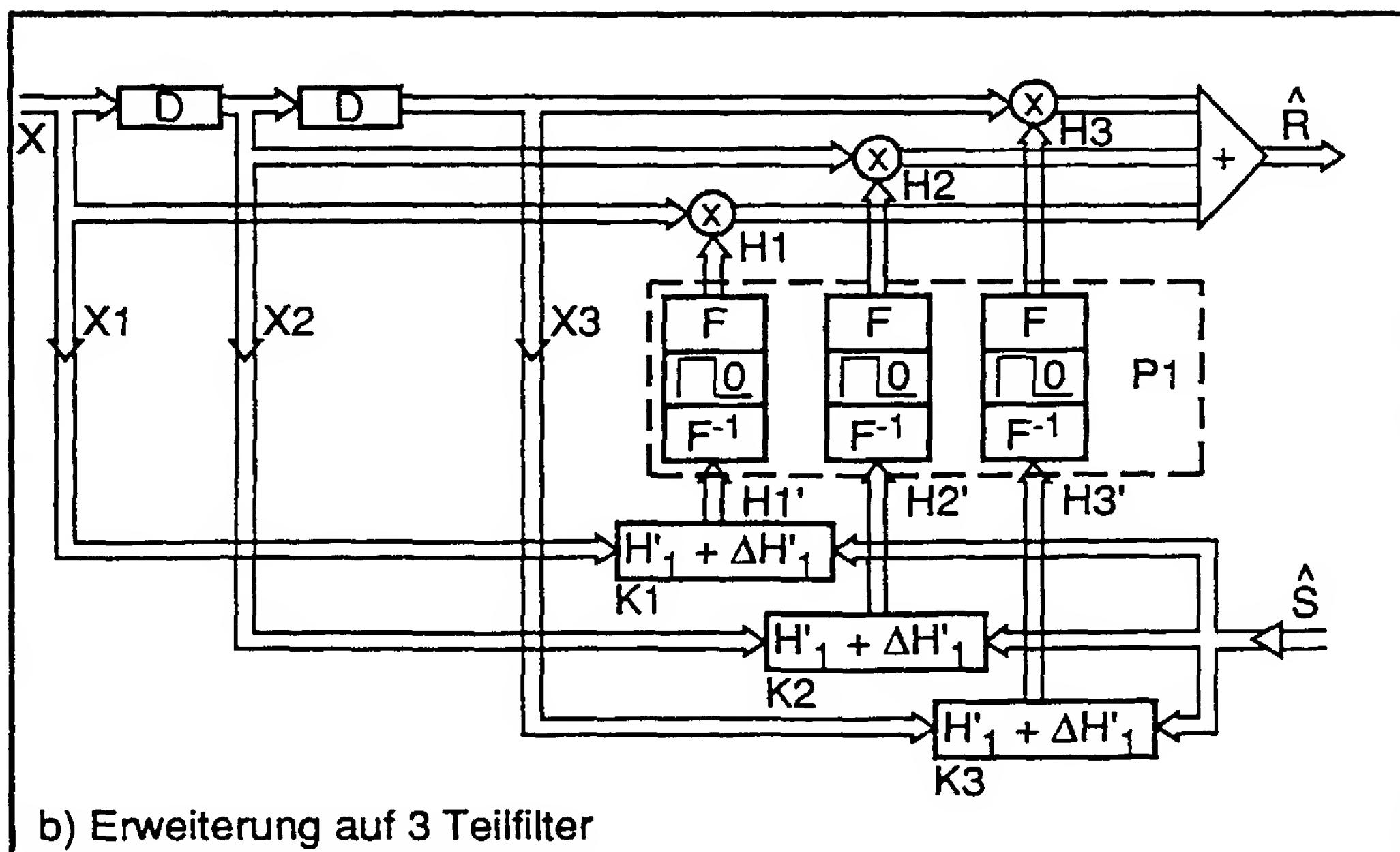
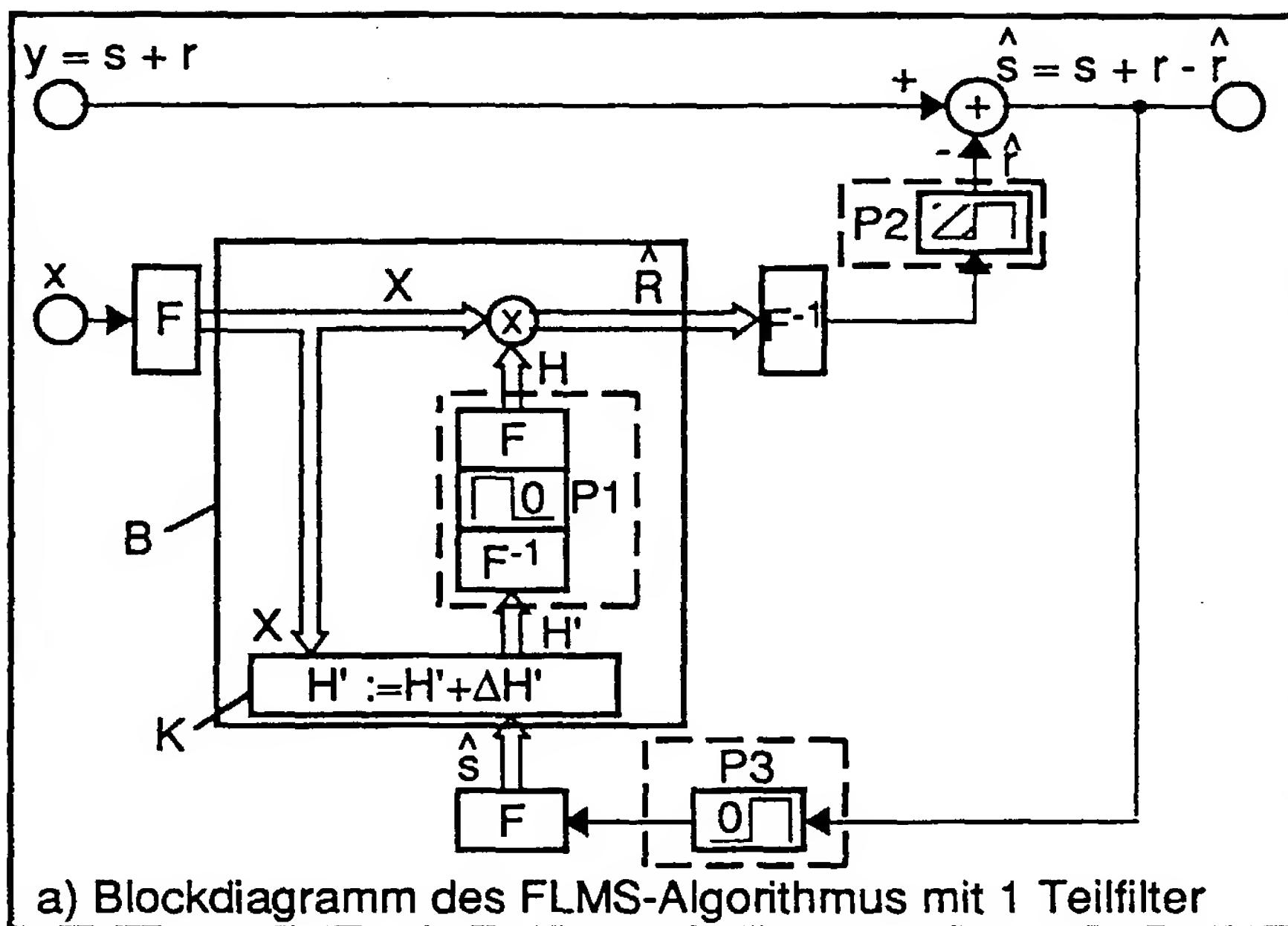


FIG.4

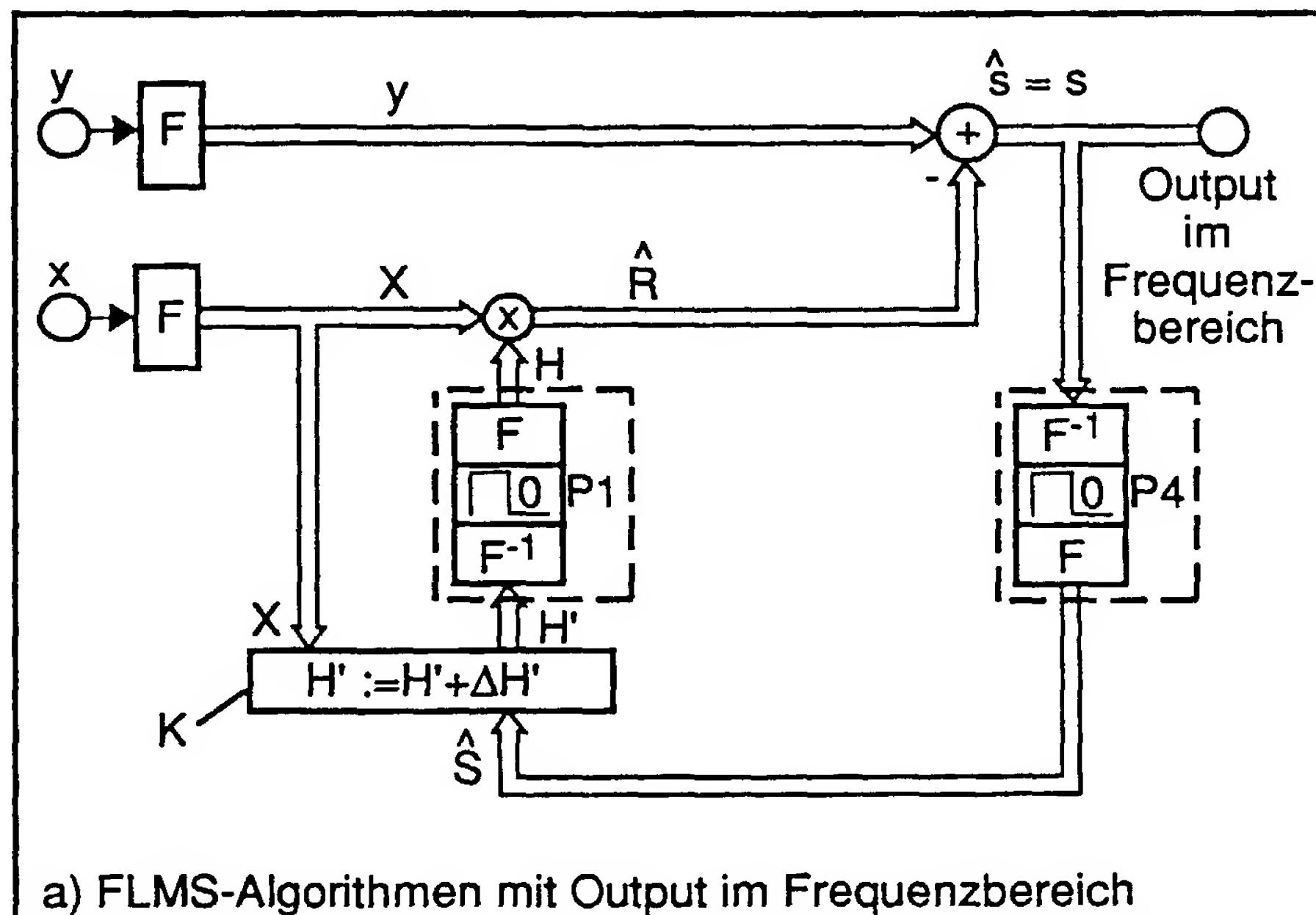
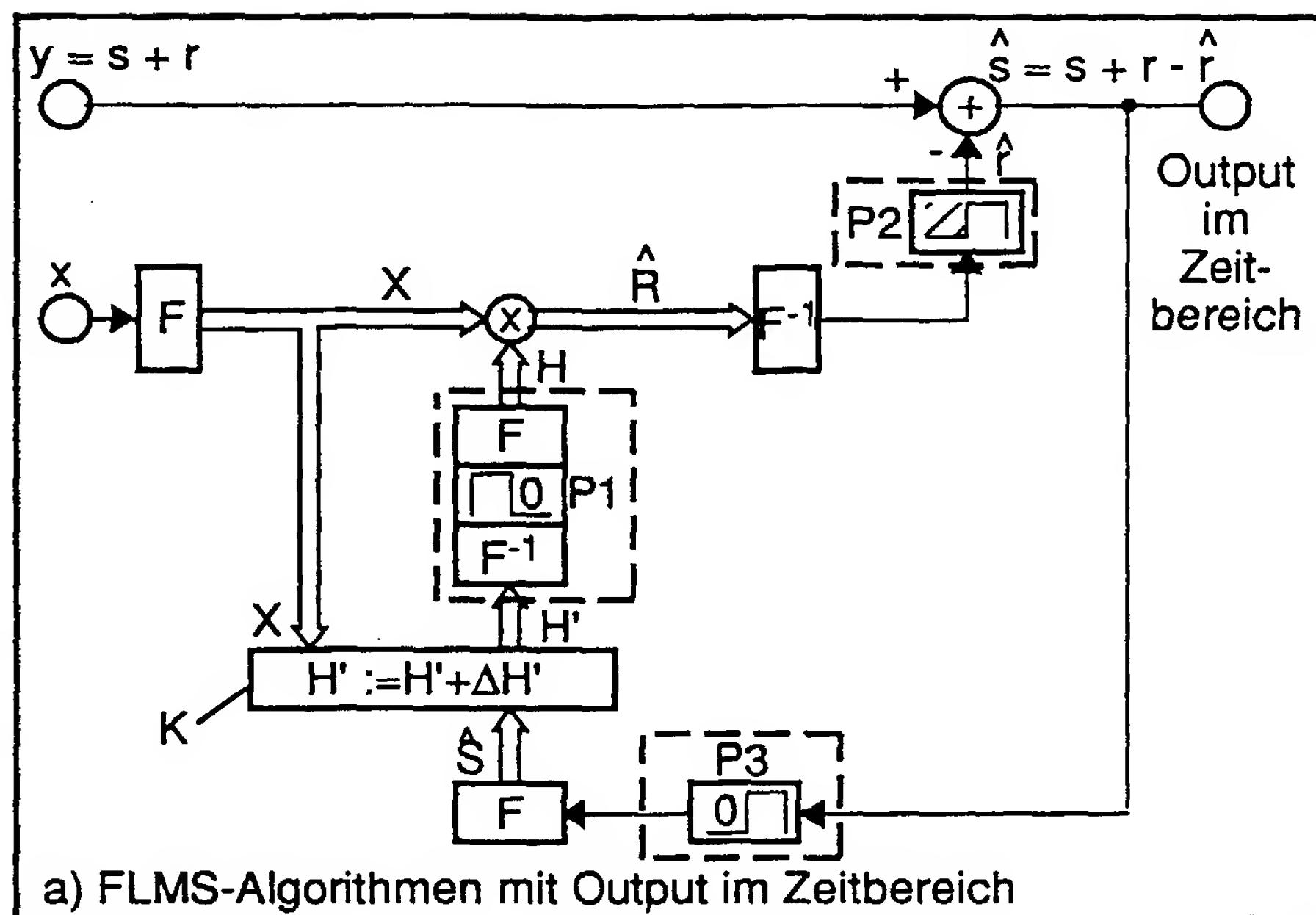


FIG.5

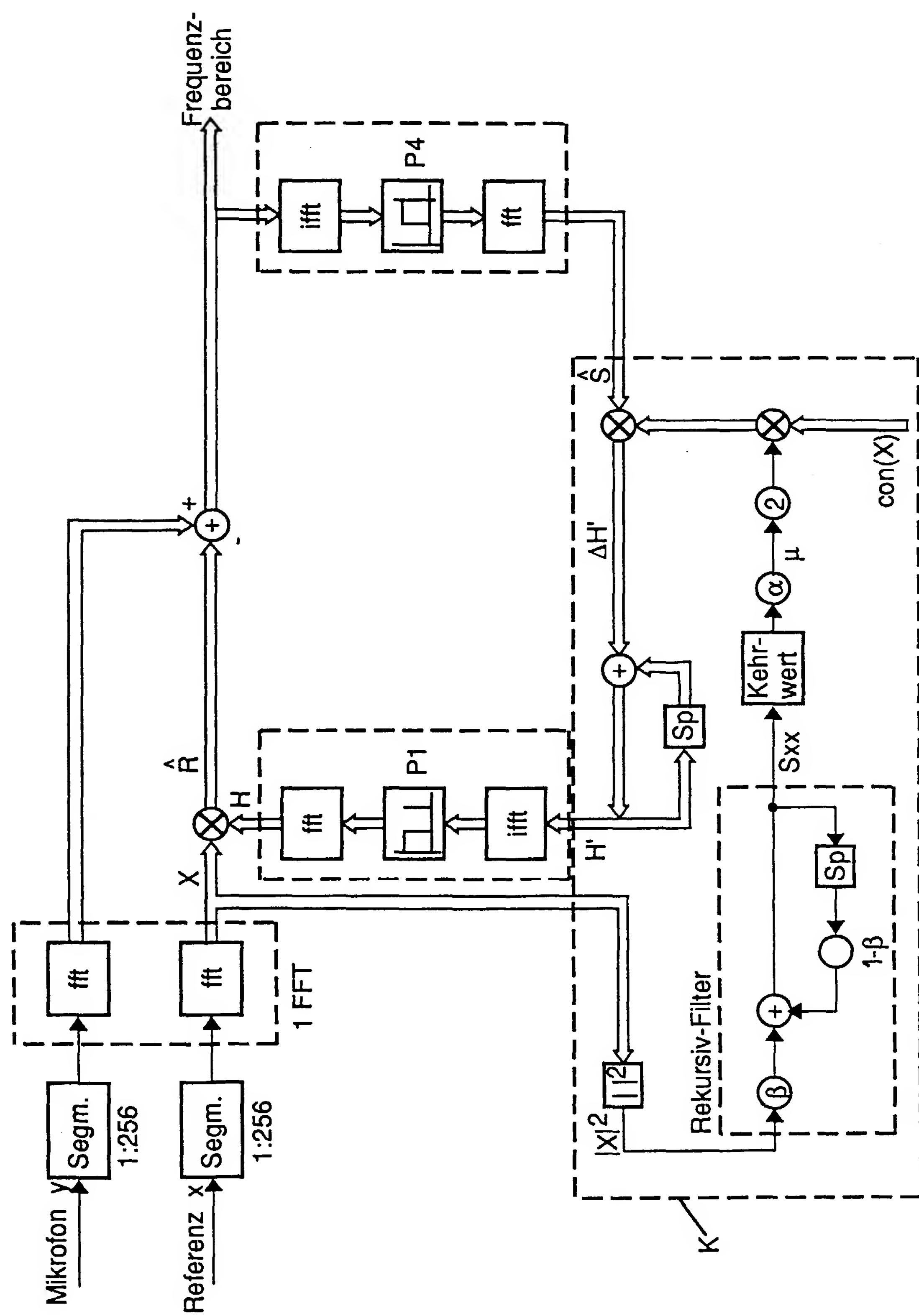


FIG.6

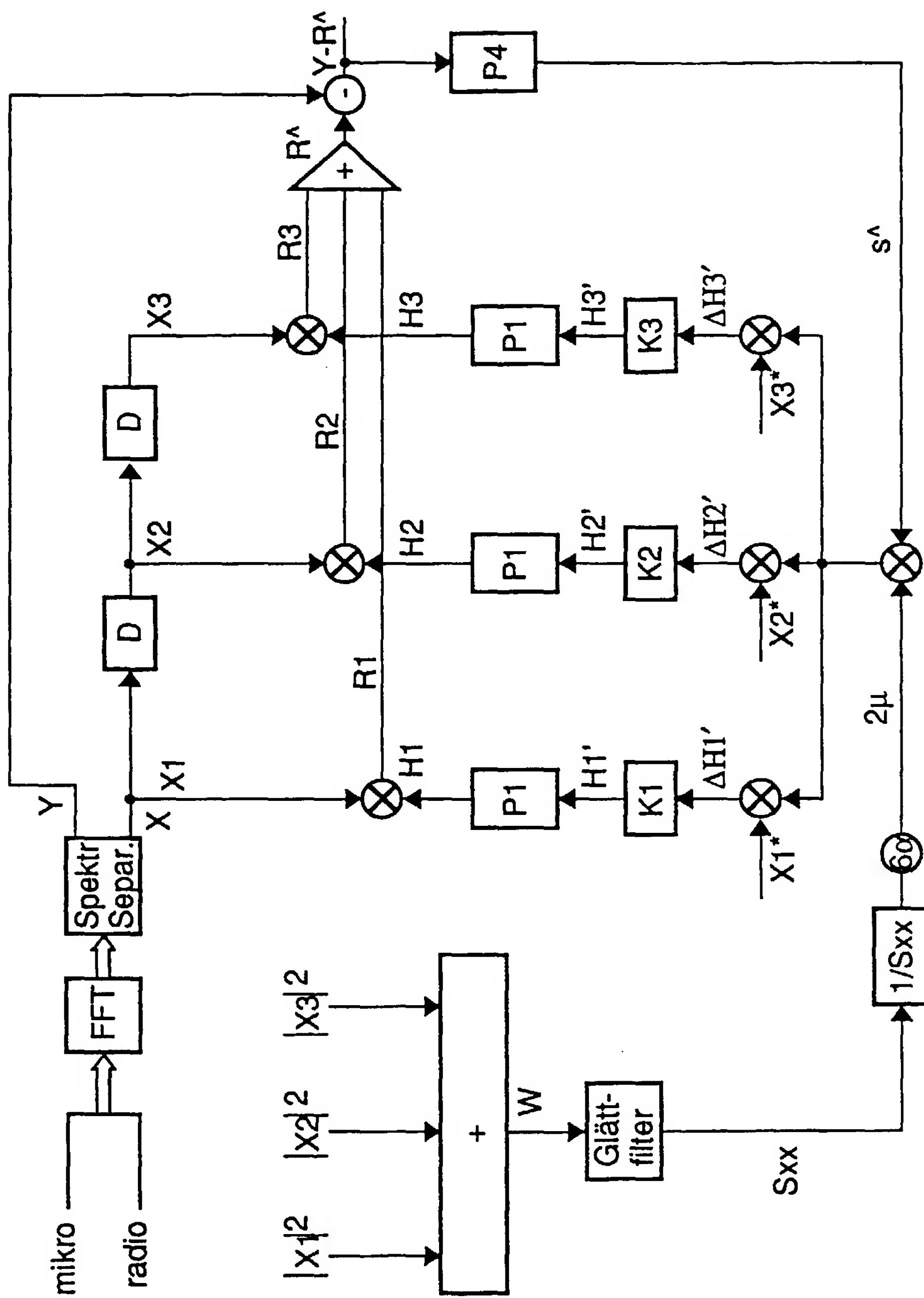


FIG.7

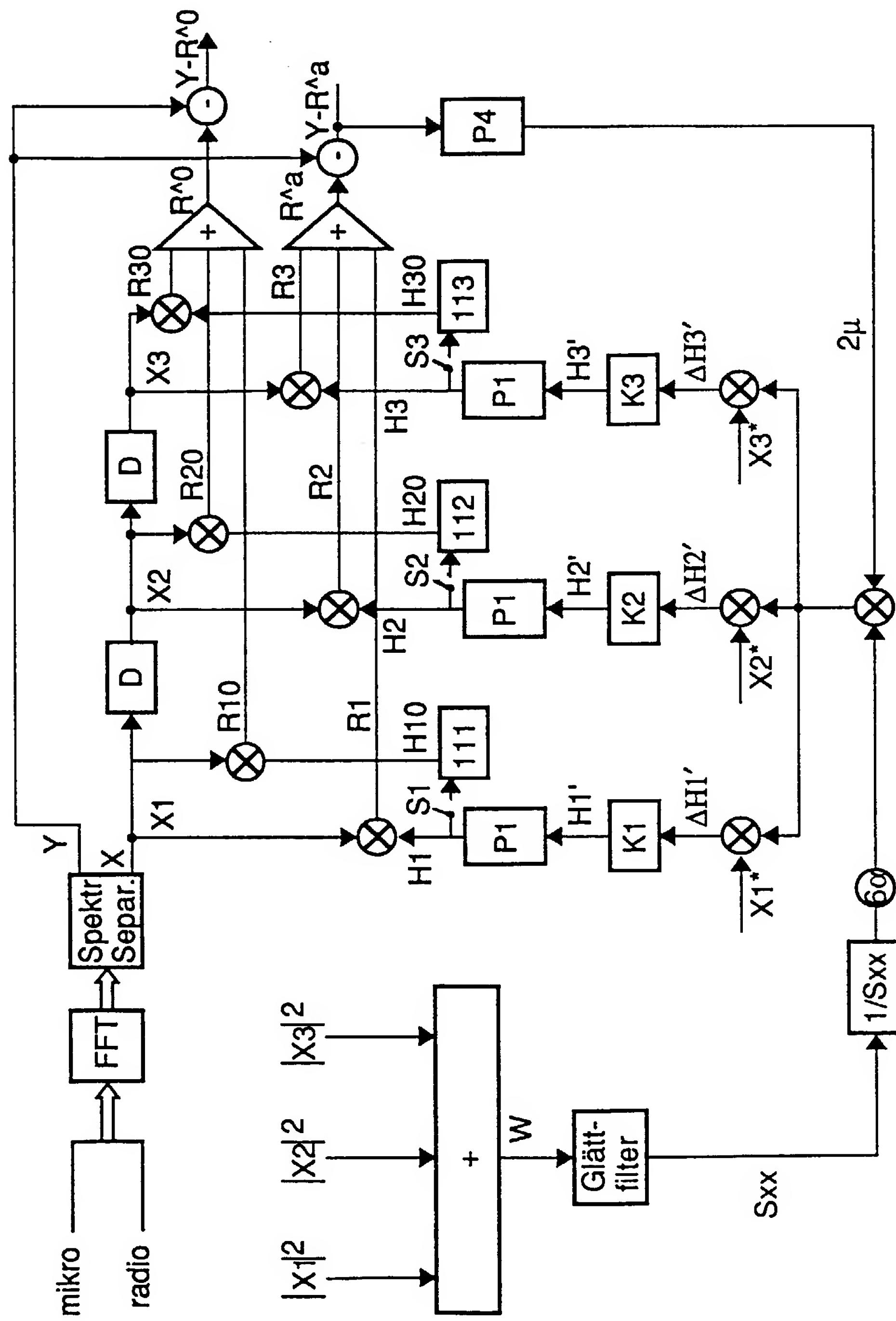


FIG.8

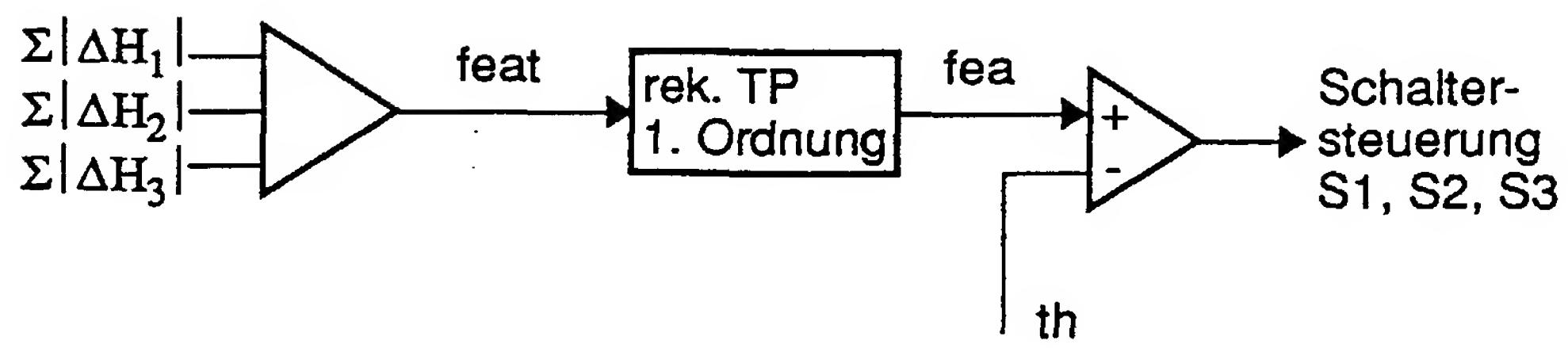


FIG.8a

Szene Z0

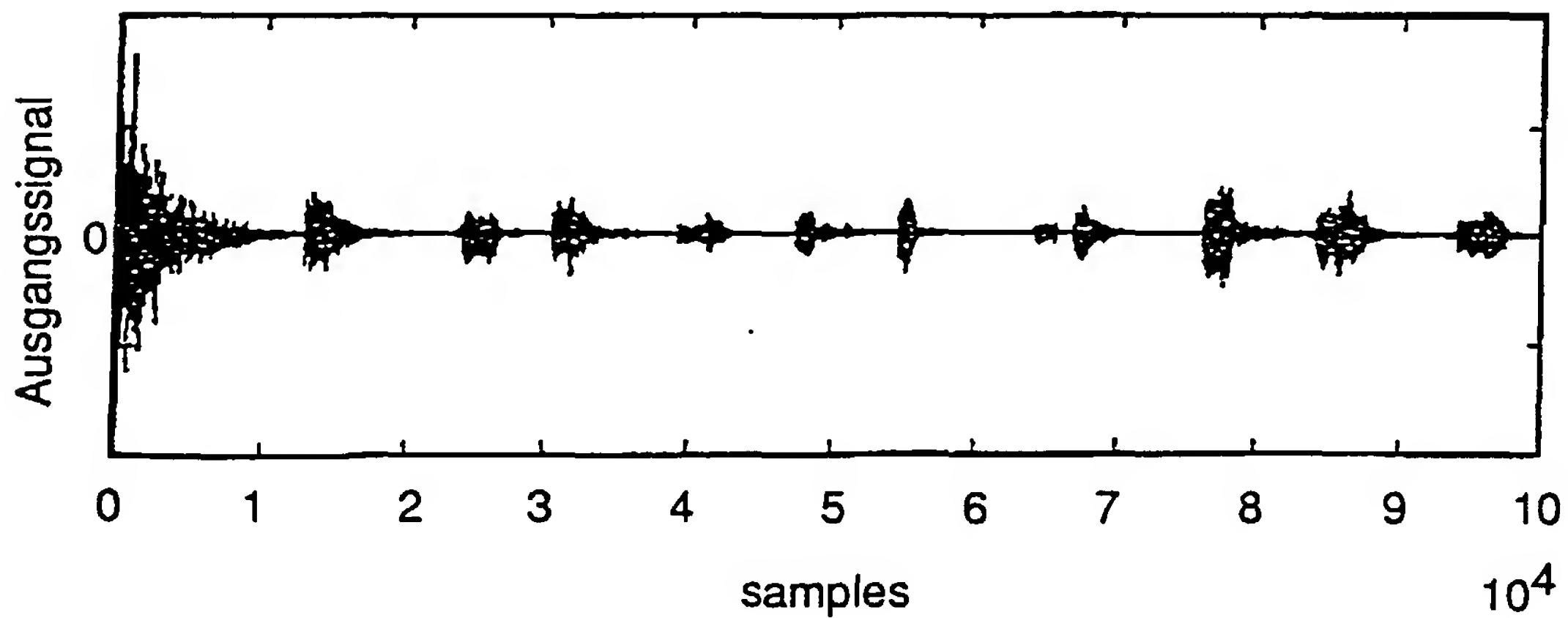
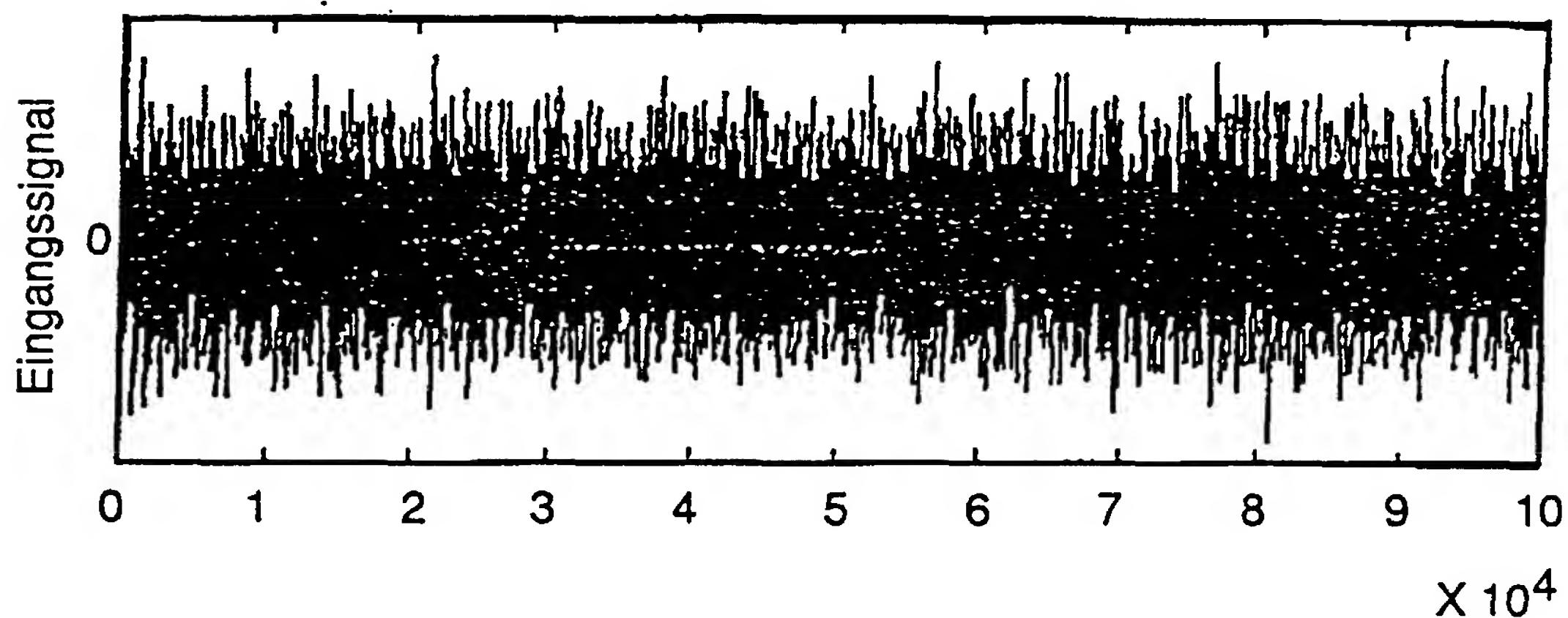


FIG.9

Szene Z0

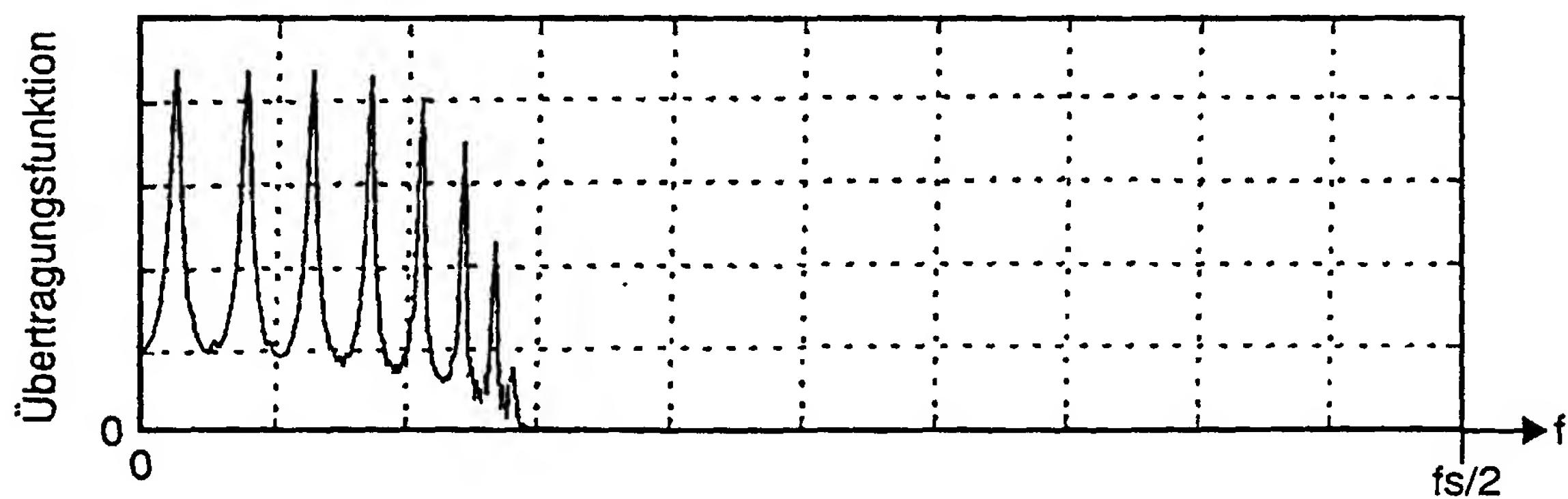
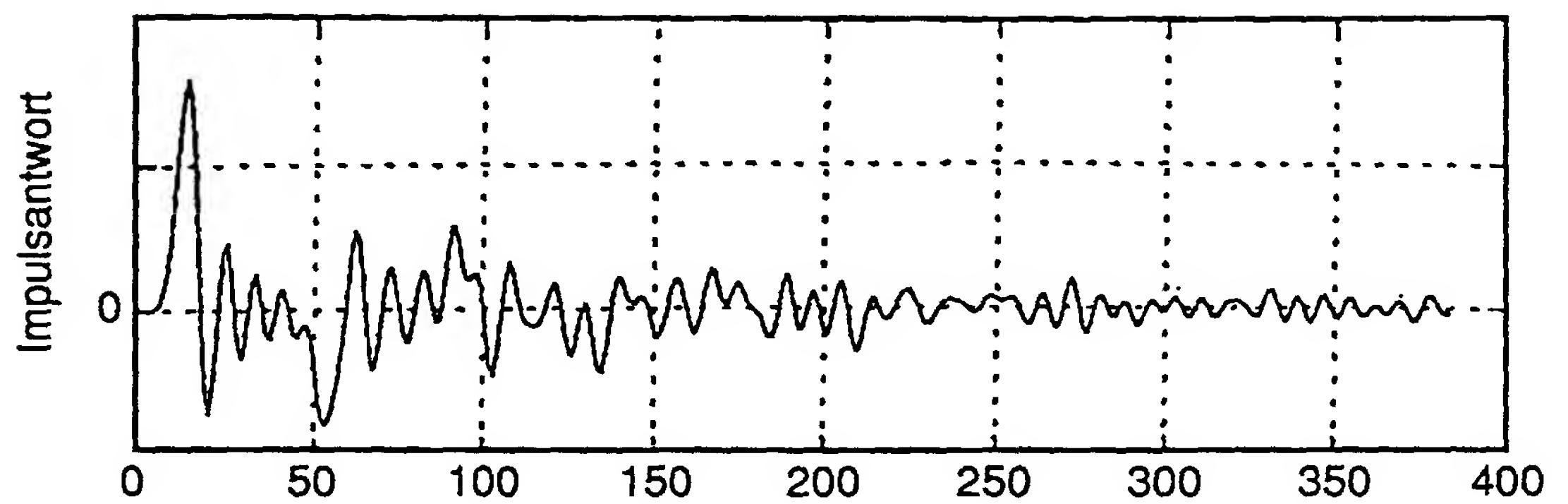


FIG.10

Szene Z1/1

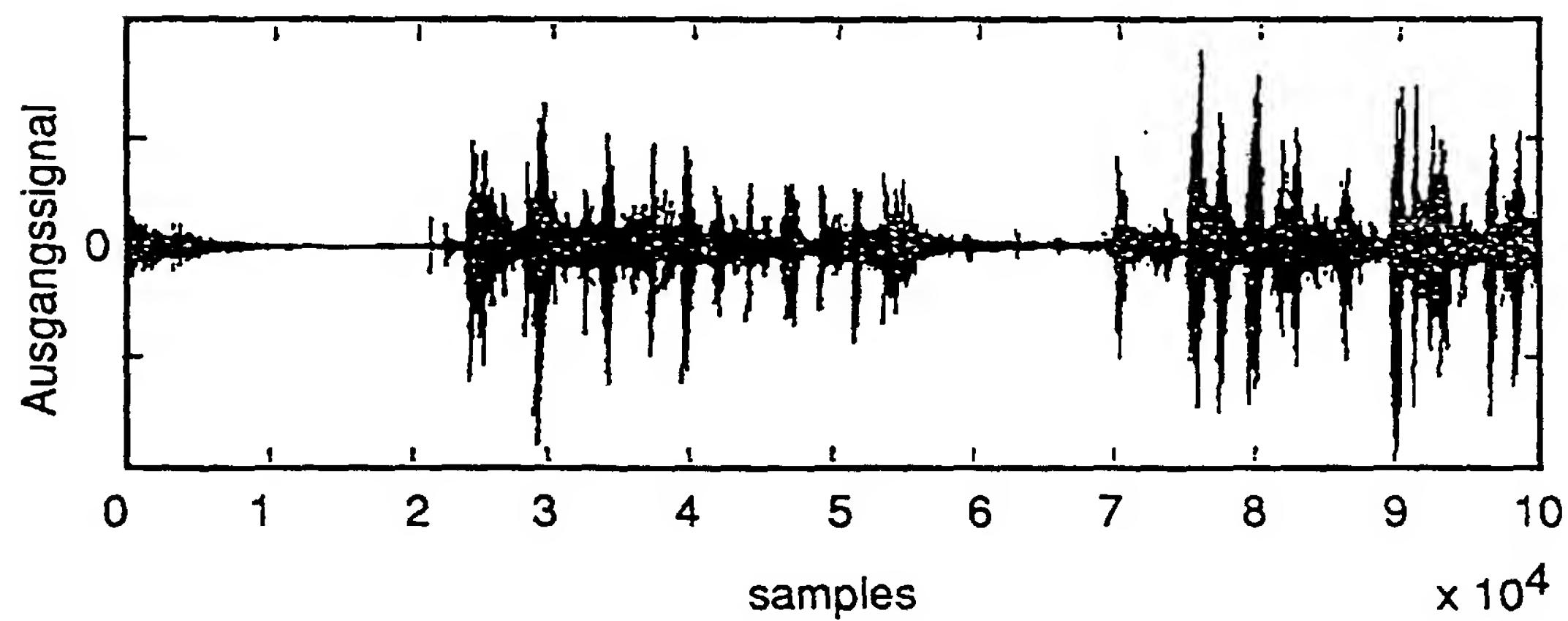
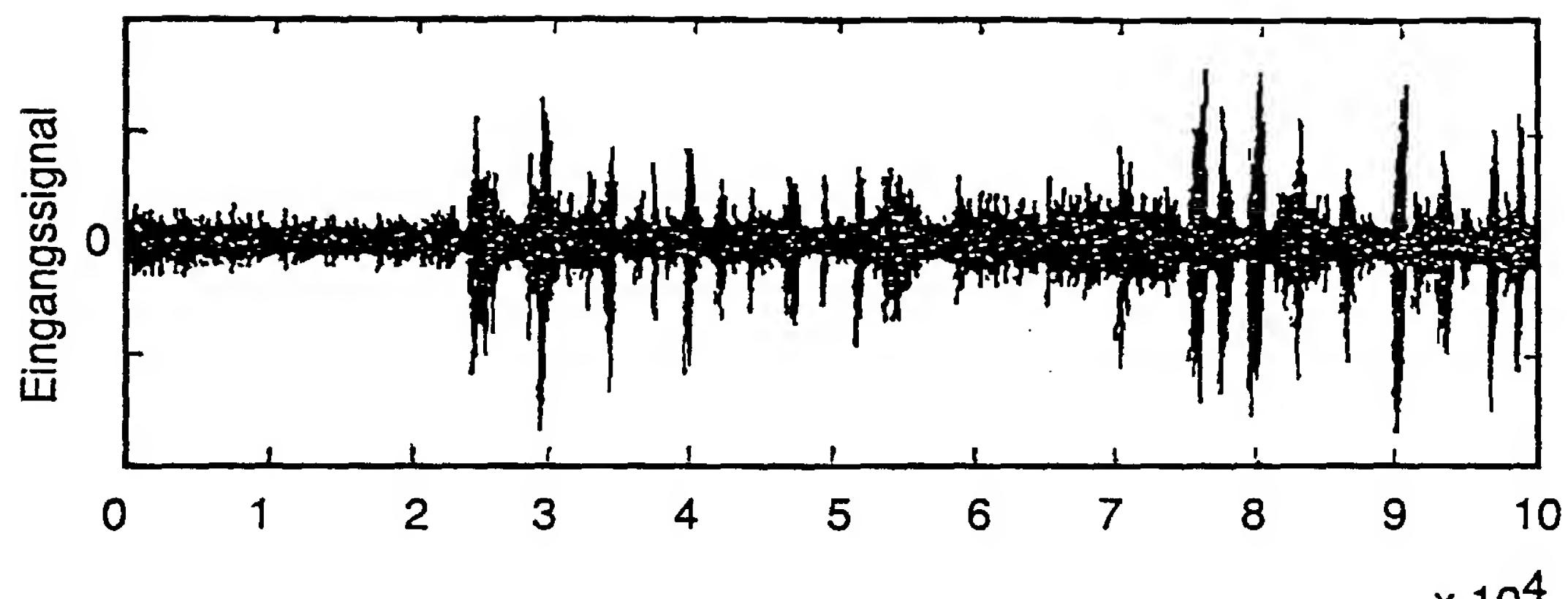


FIG.11

Szene Z1/1

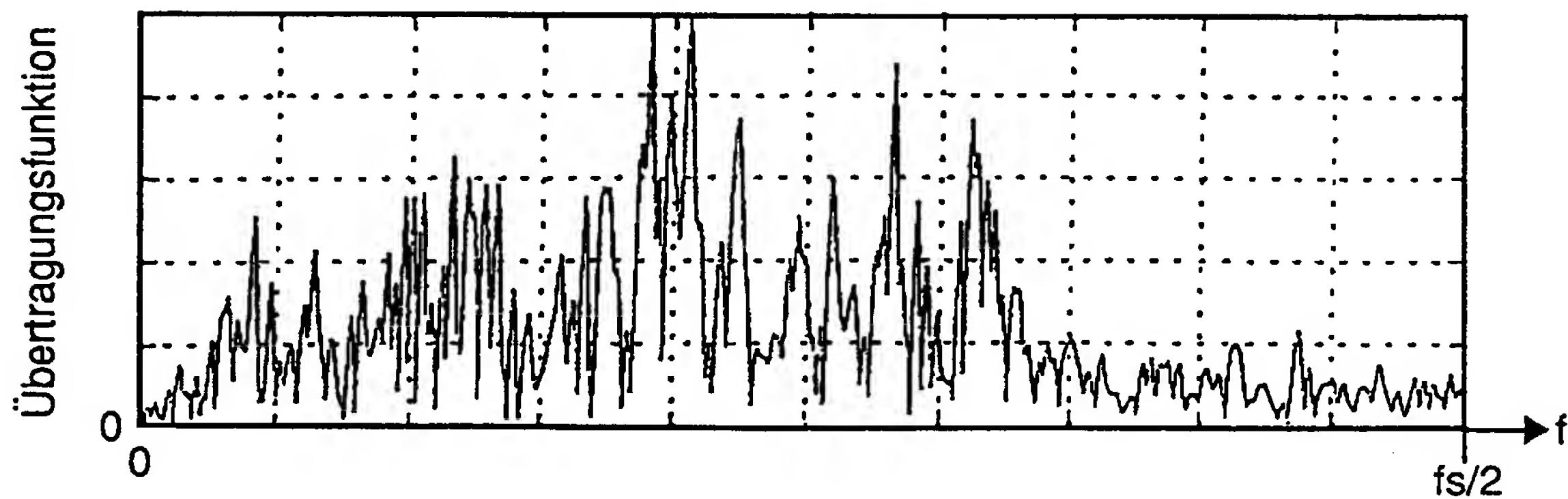
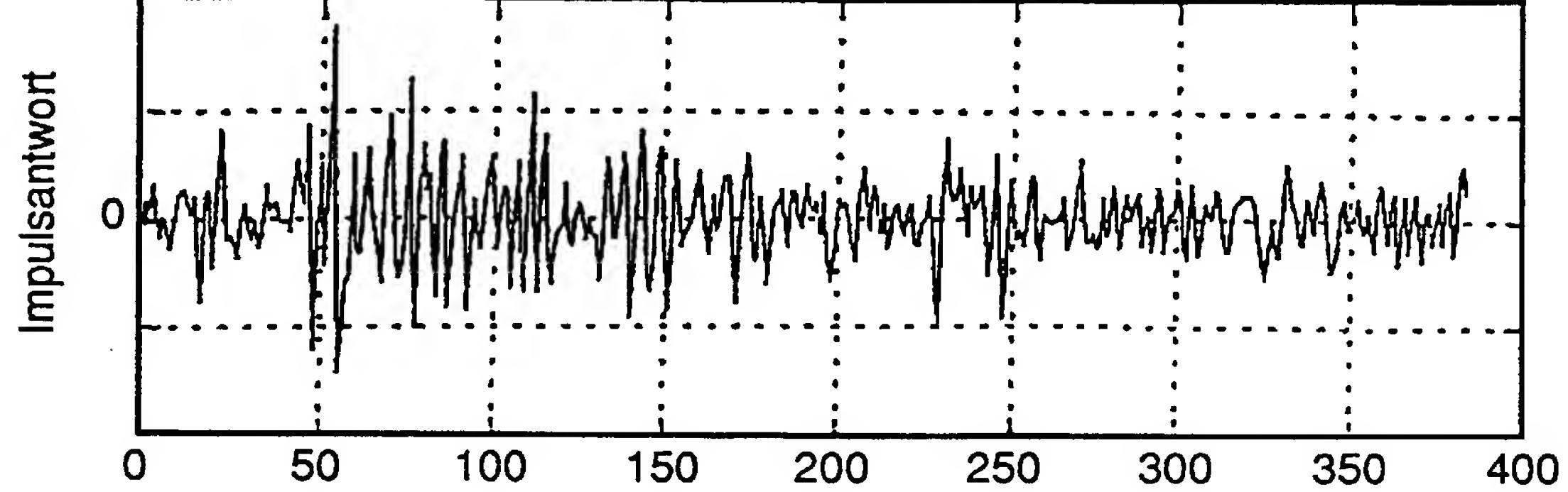


FIG.12

Szene Z1/2

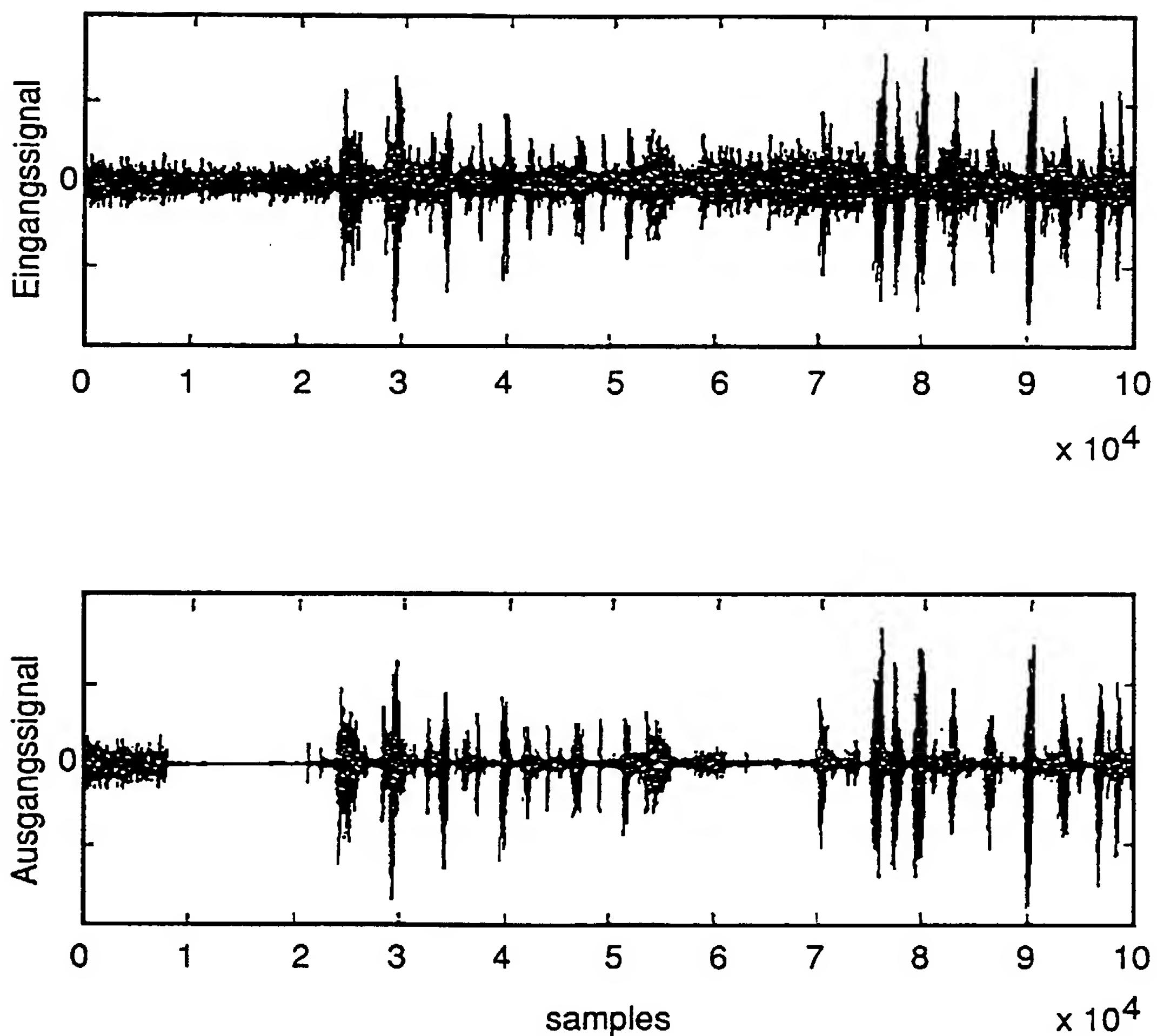


FIG.13

Szene Z1/2

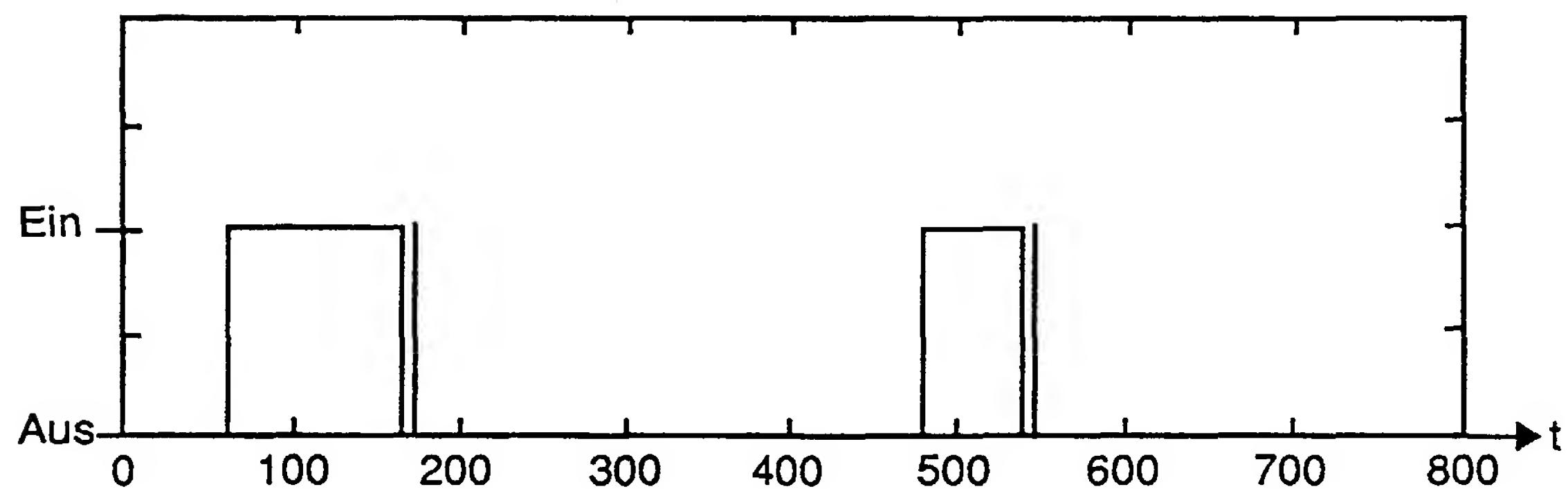
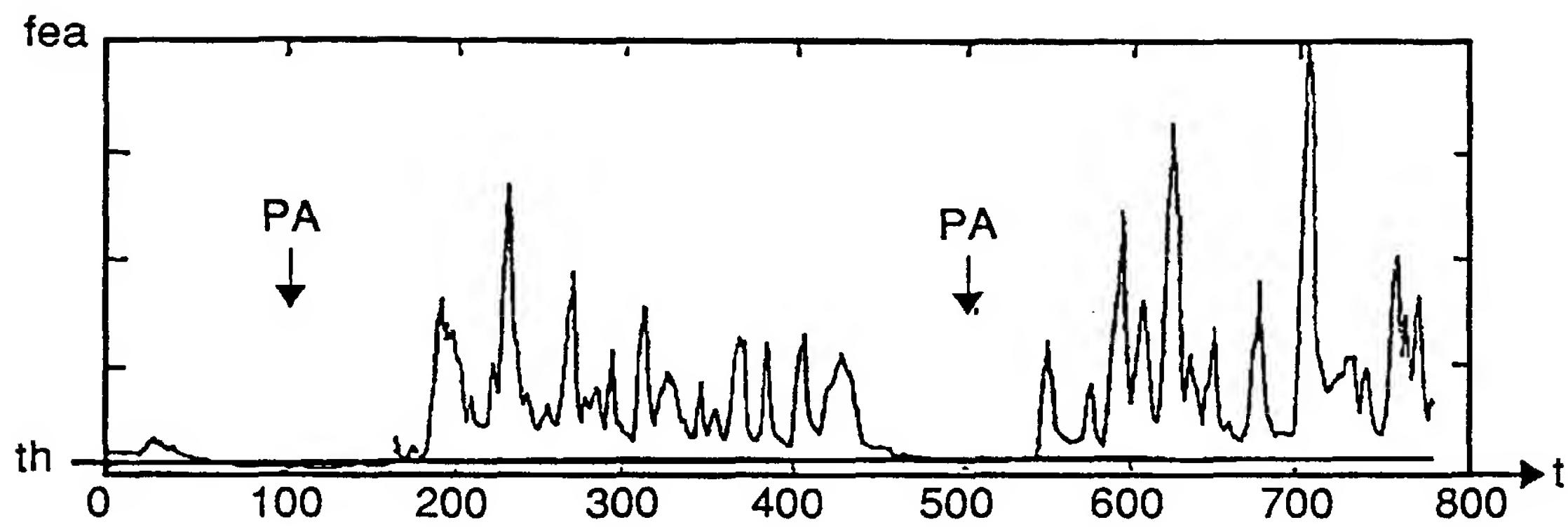


FIG.14

Szene Z1

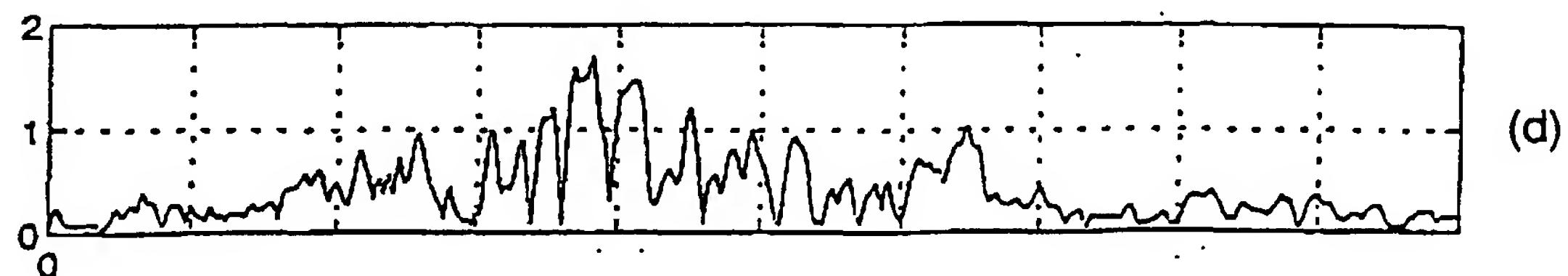
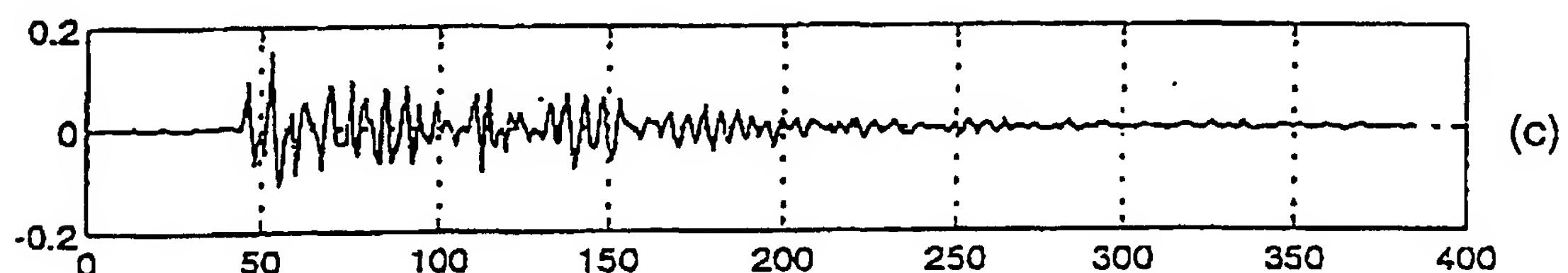
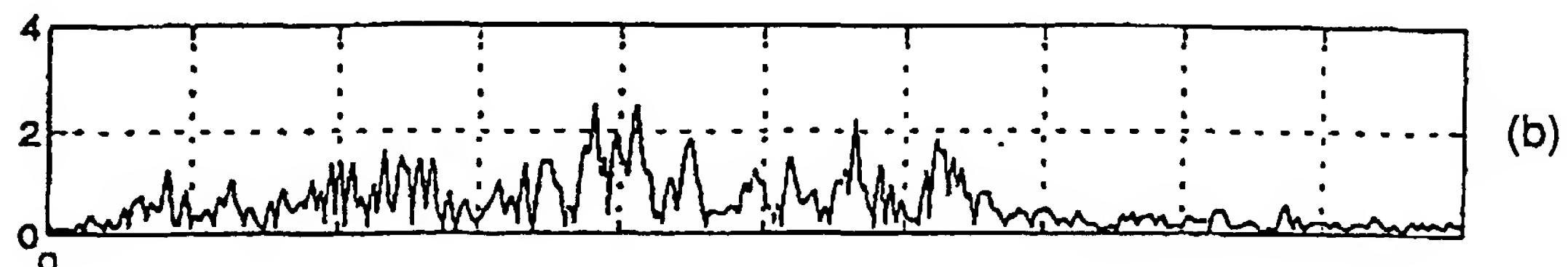
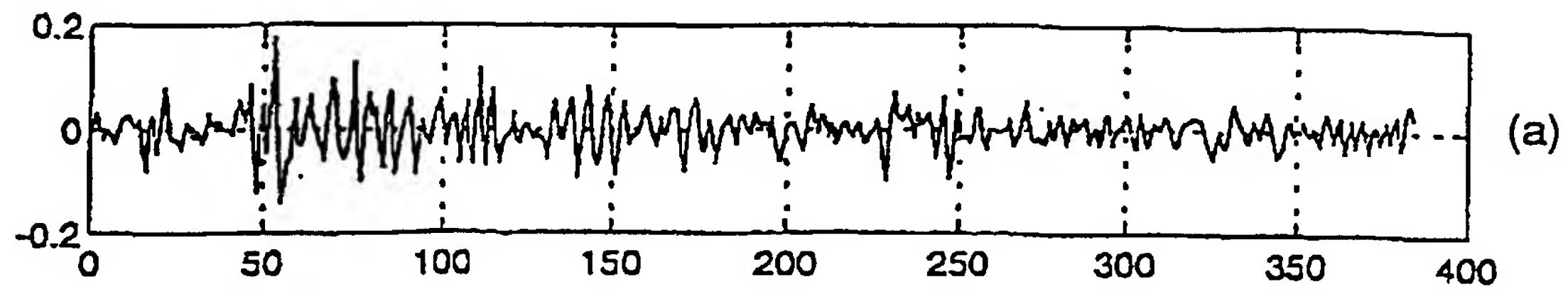
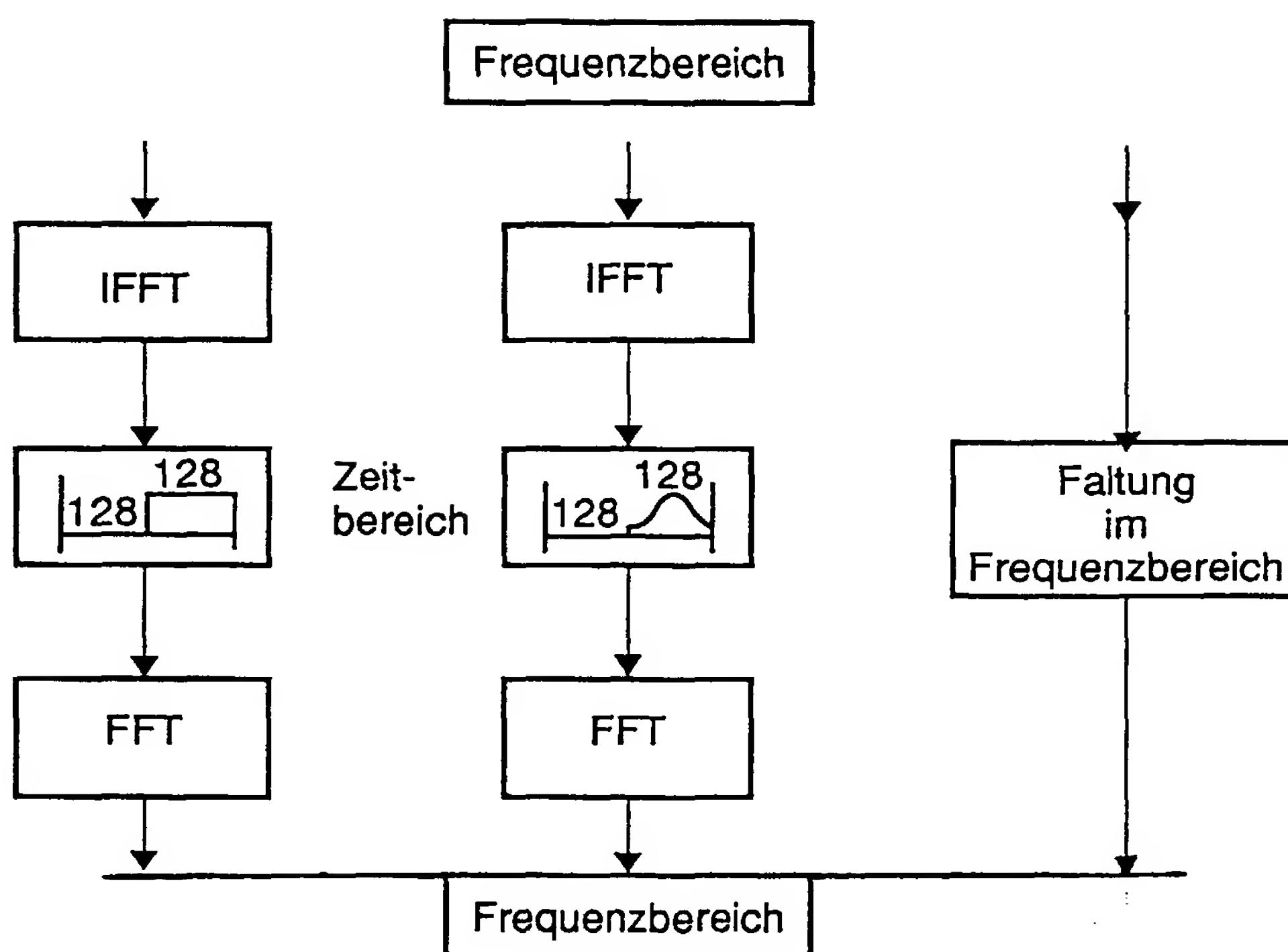


FIG.15

a)

b)

c)



ideale Projektion P4
mit Rechteck-Fenster
im Zeitbereich

modifizierte Projektion P4
mit Hamming-Fenster
im Zeitbereich

zu b) äquivalente Faltung
im Frequenzbereich (Faltung
mit 7-Linien Spektrum)

FIG.16

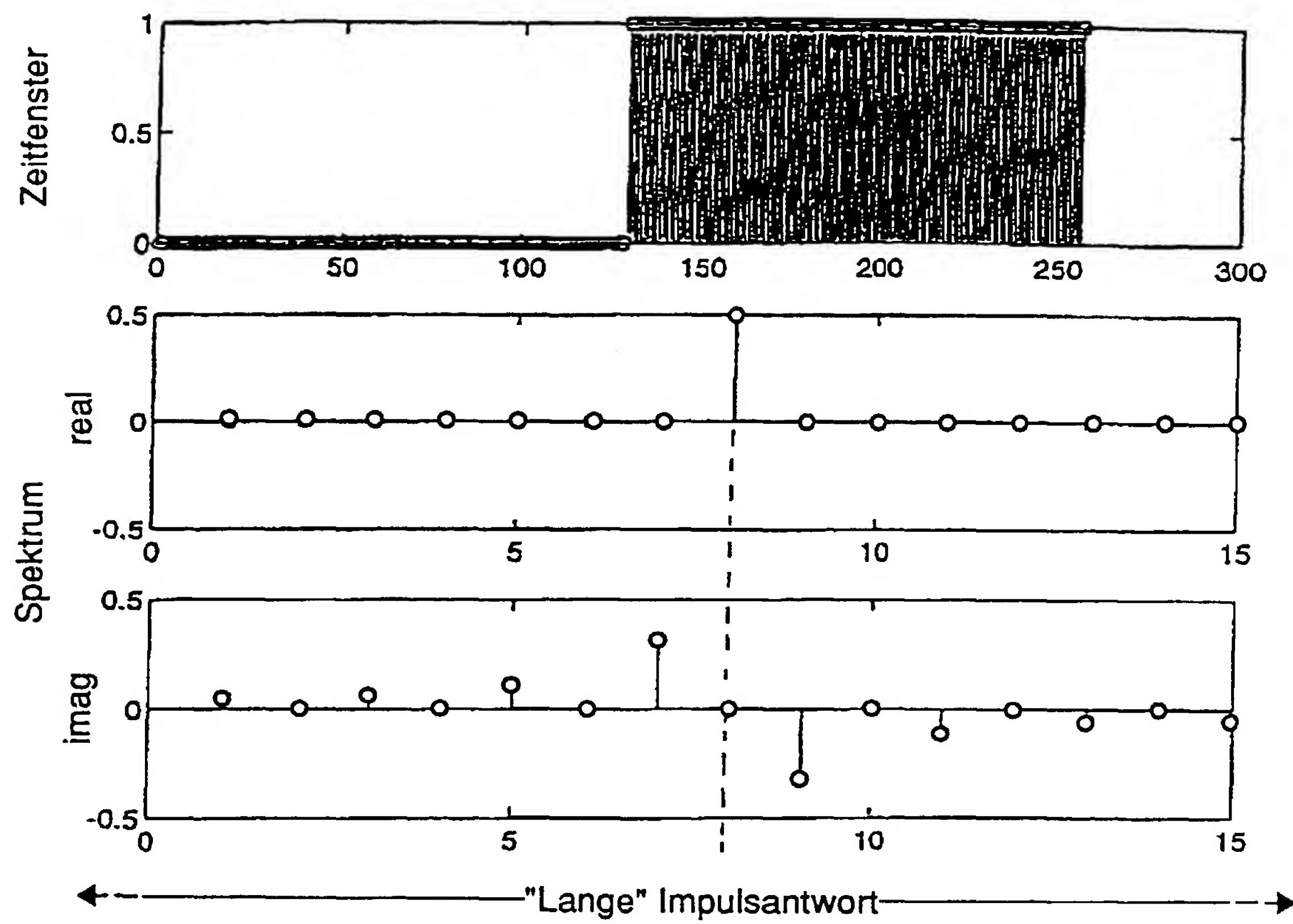


FIG.17

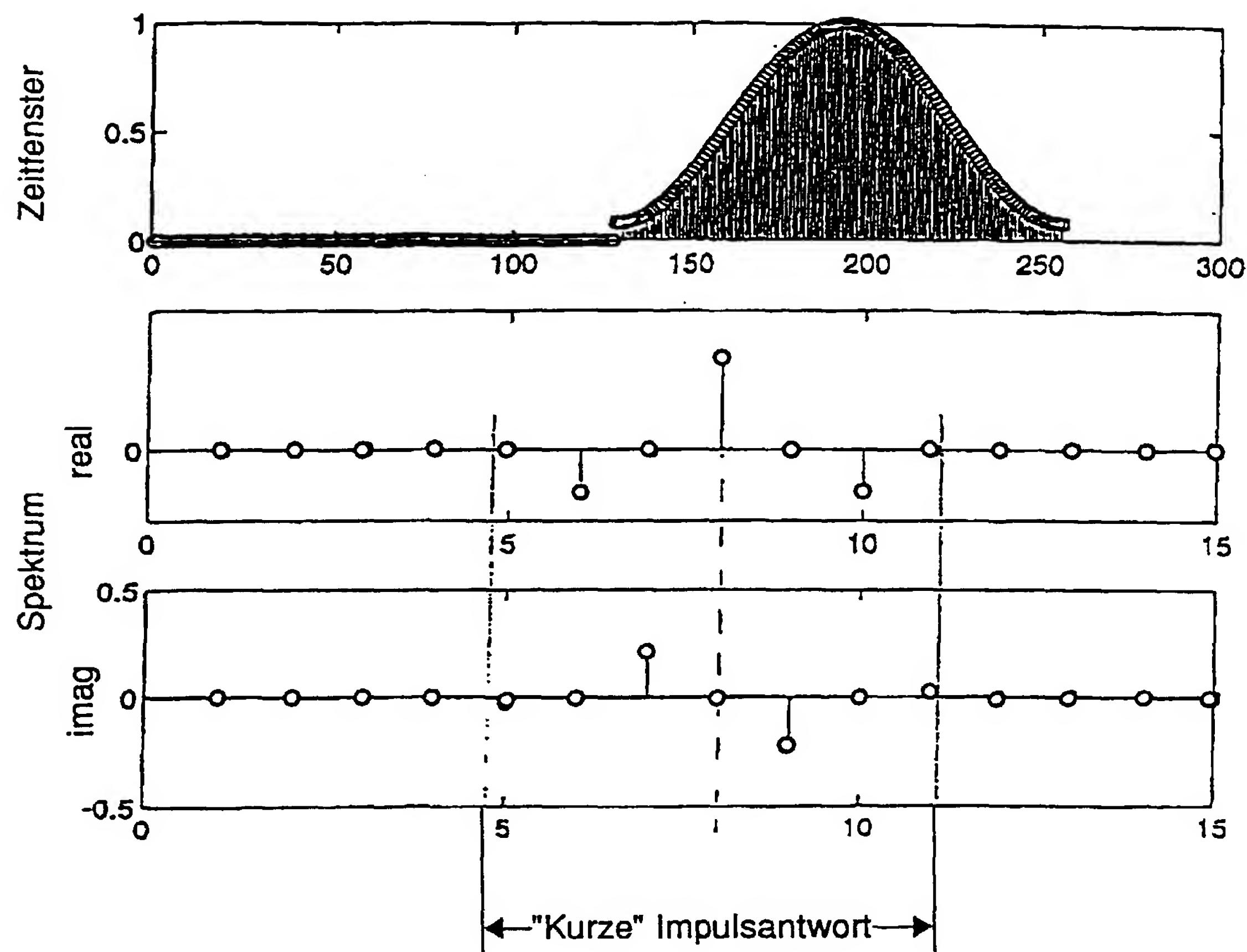


FIG.18

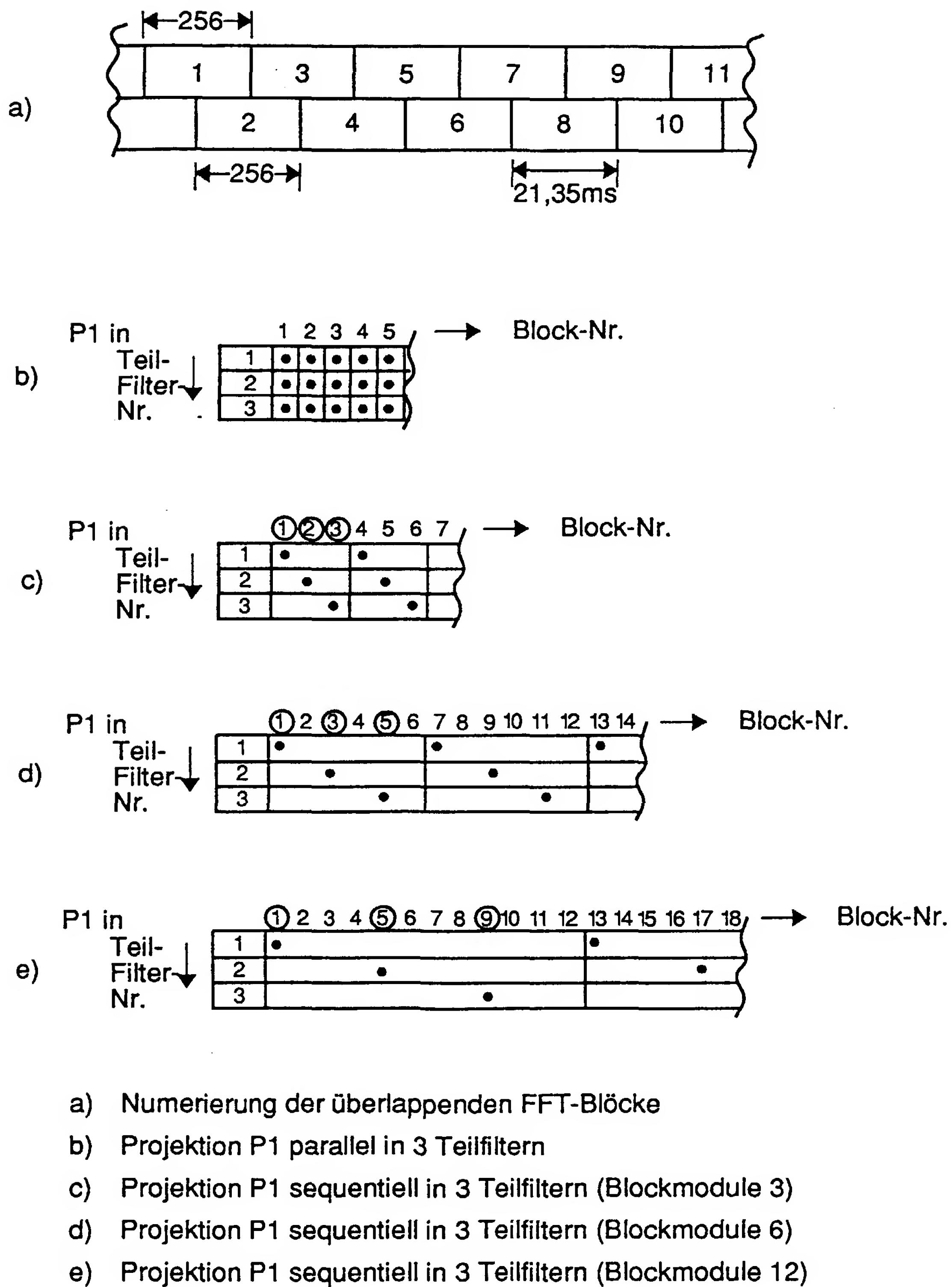


FIG.19

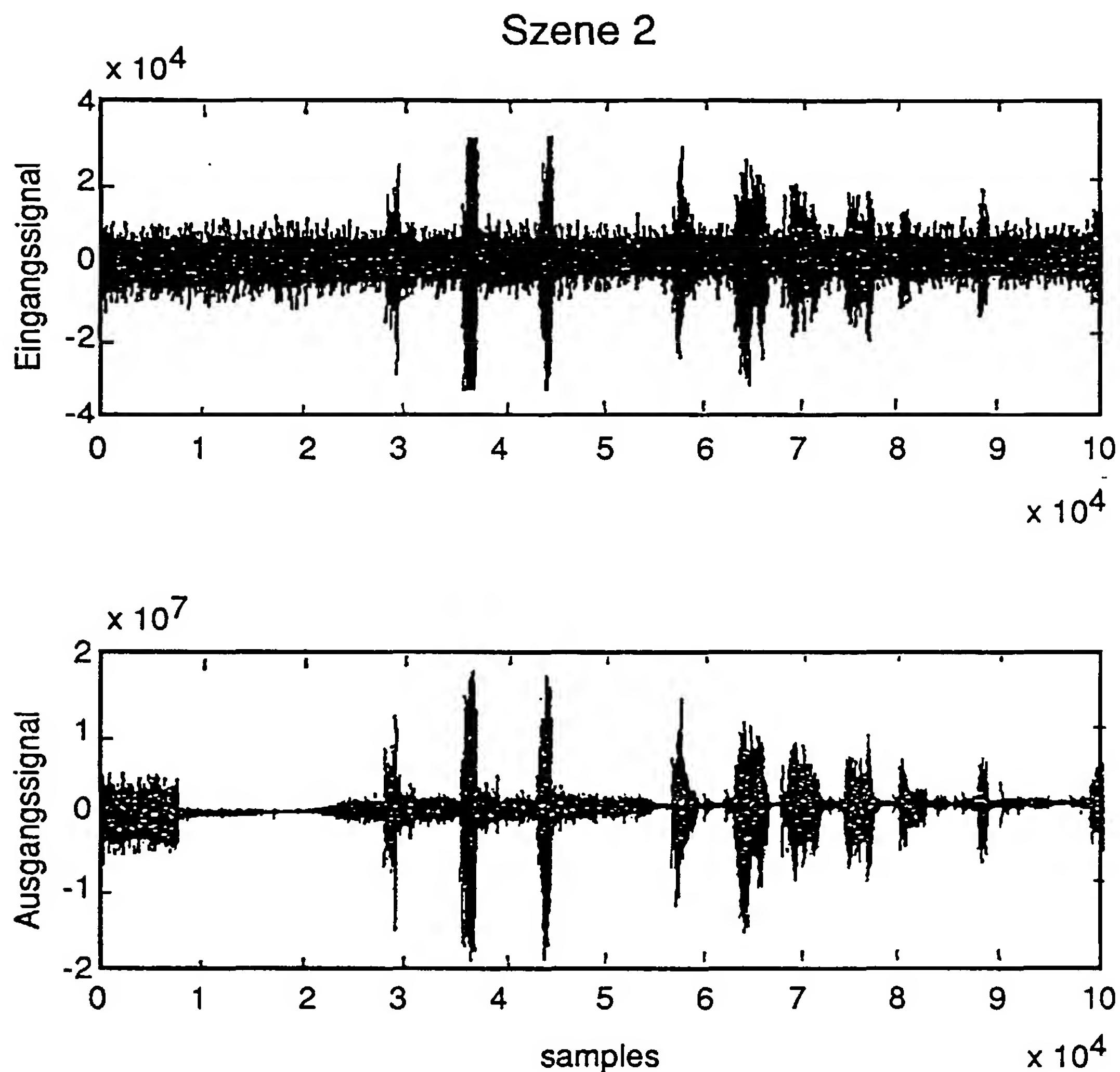


FIG.20

Szene Z2

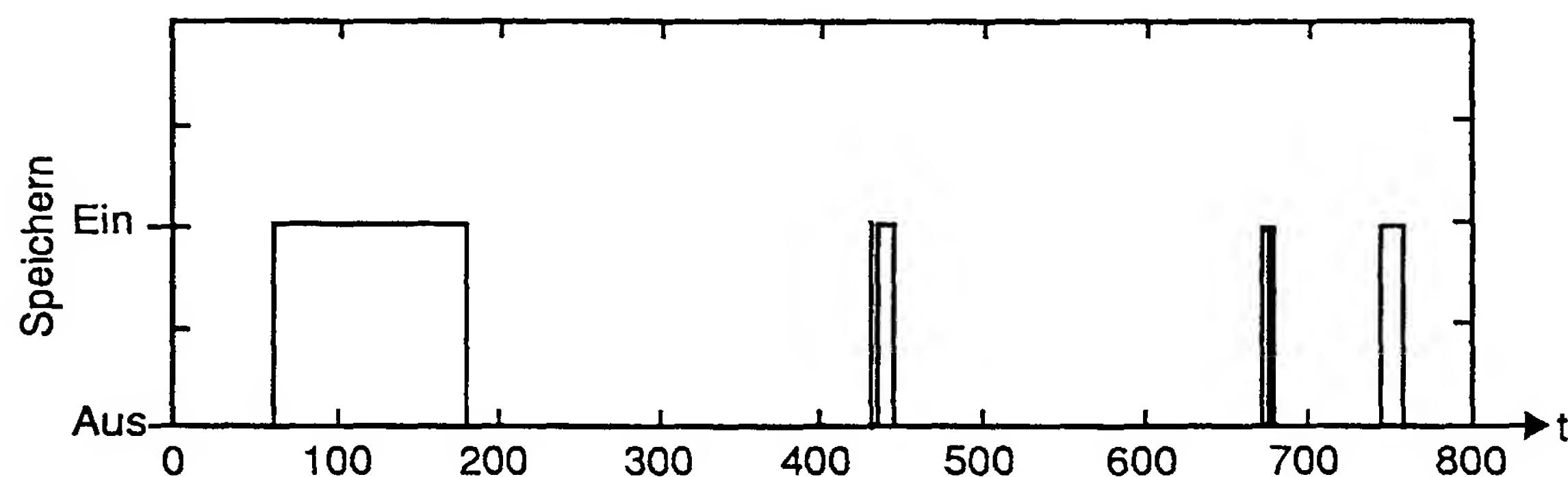
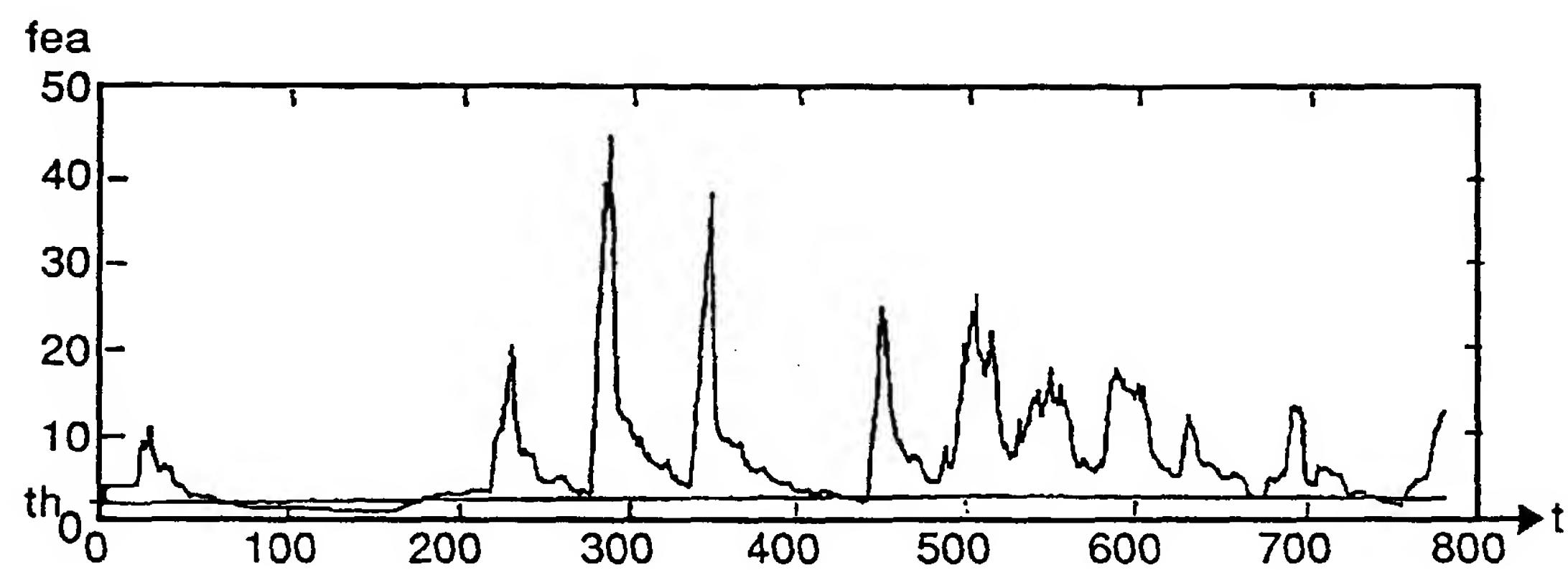


FIG.21

Szene Z2

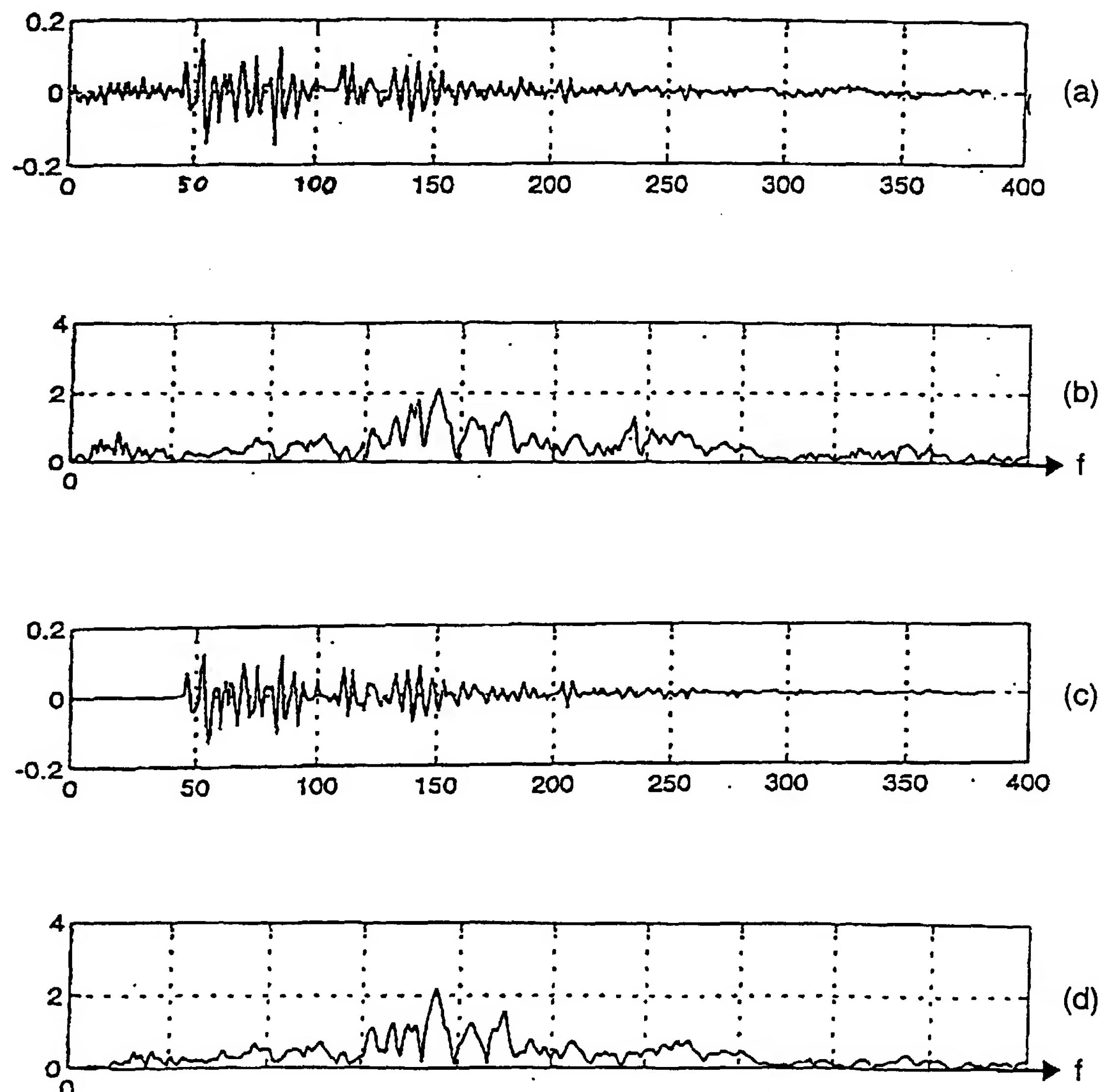


FIG.22

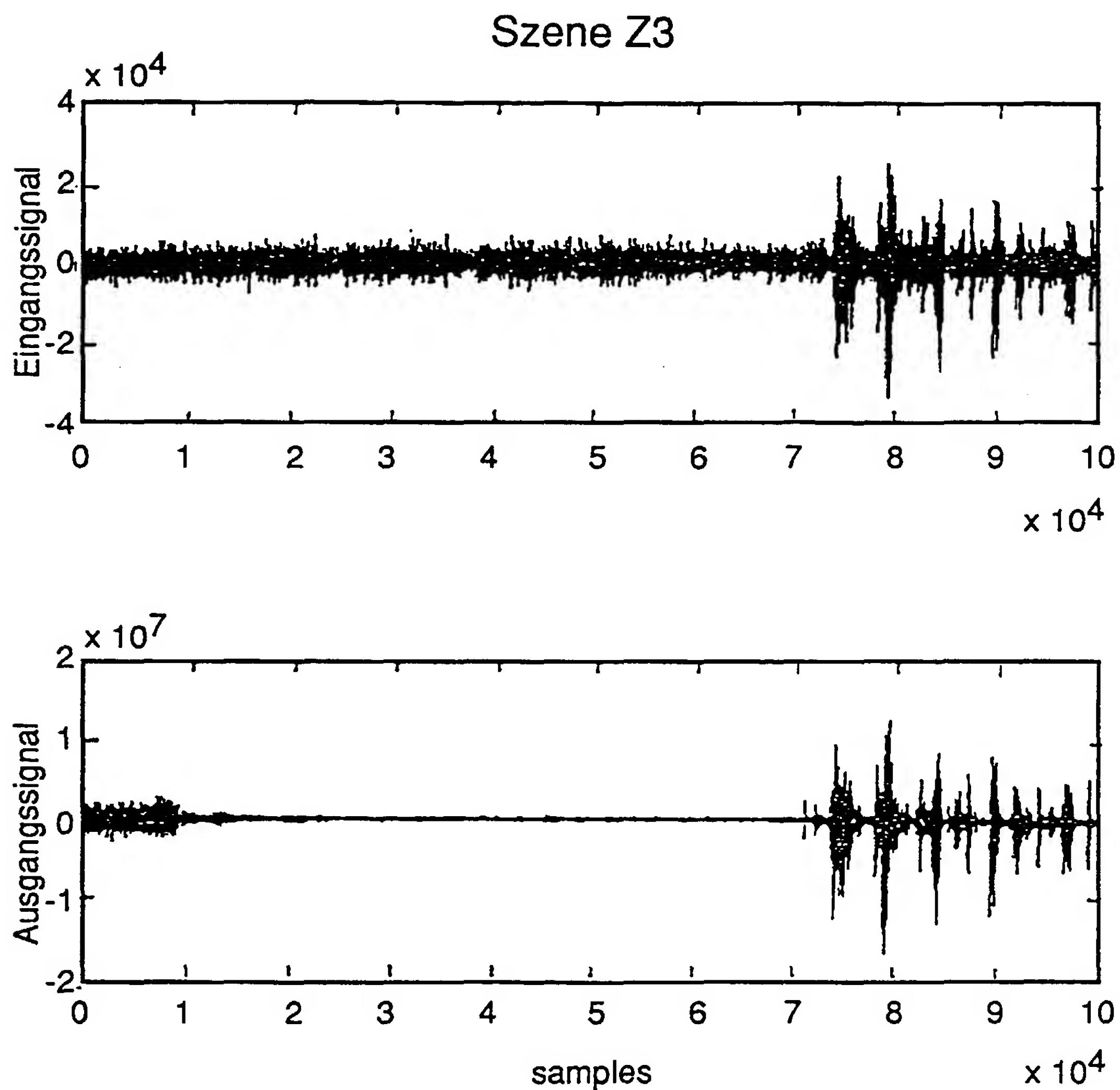


FIG.23

fea

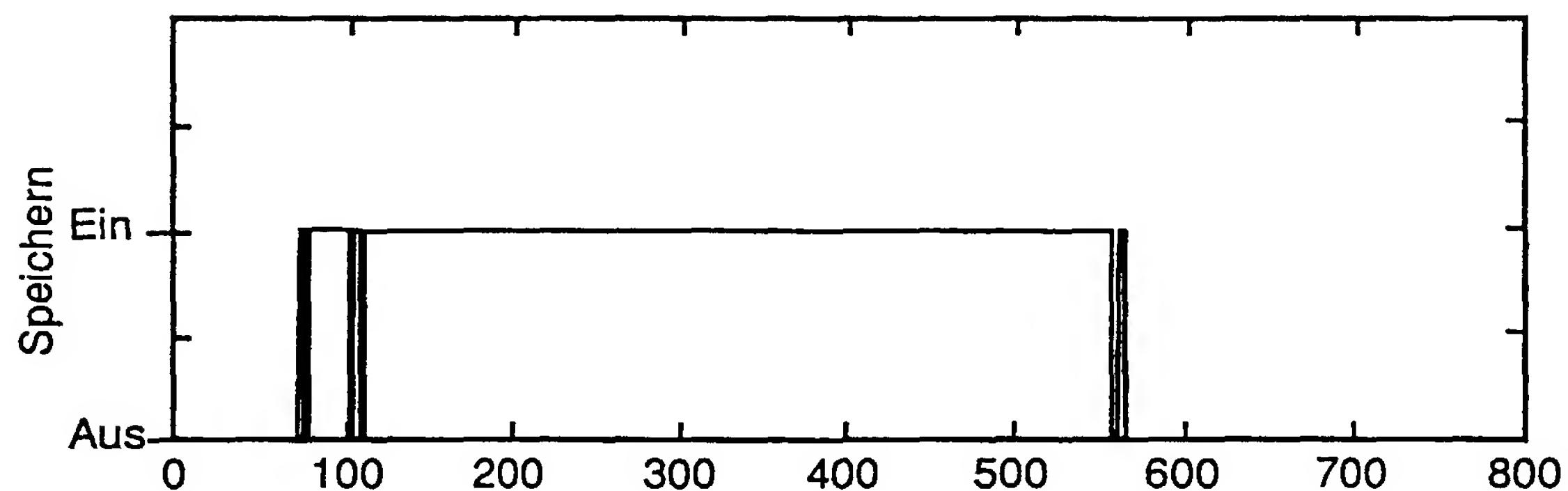
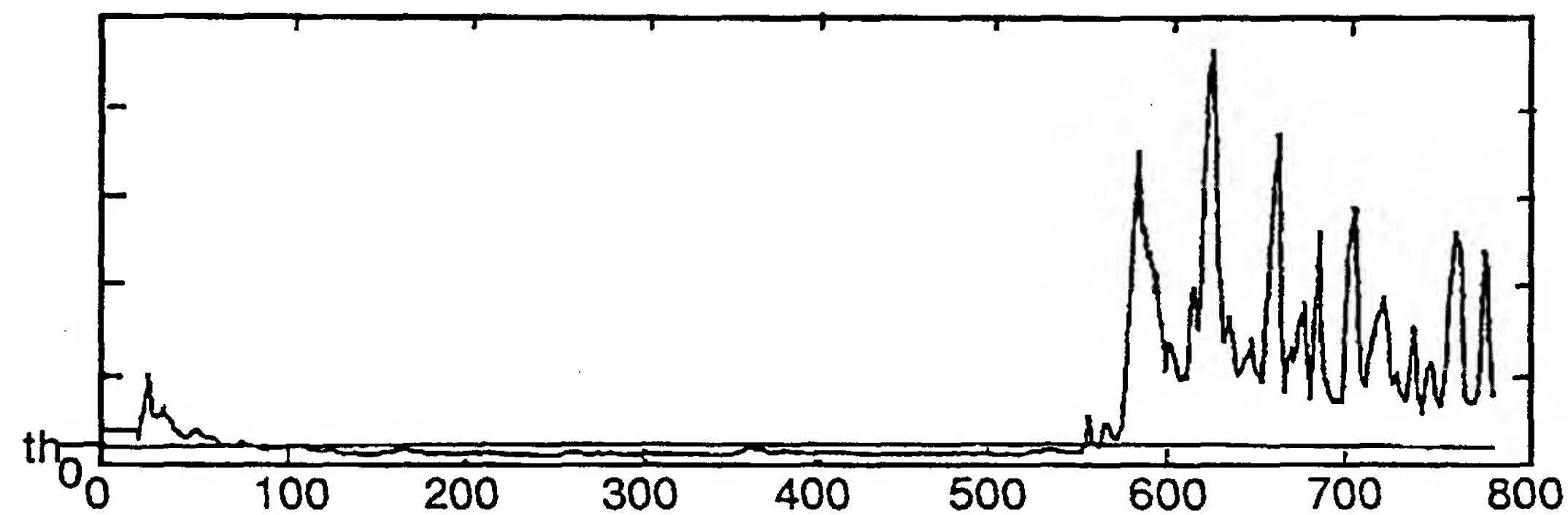


FIG.24

Szene Z3

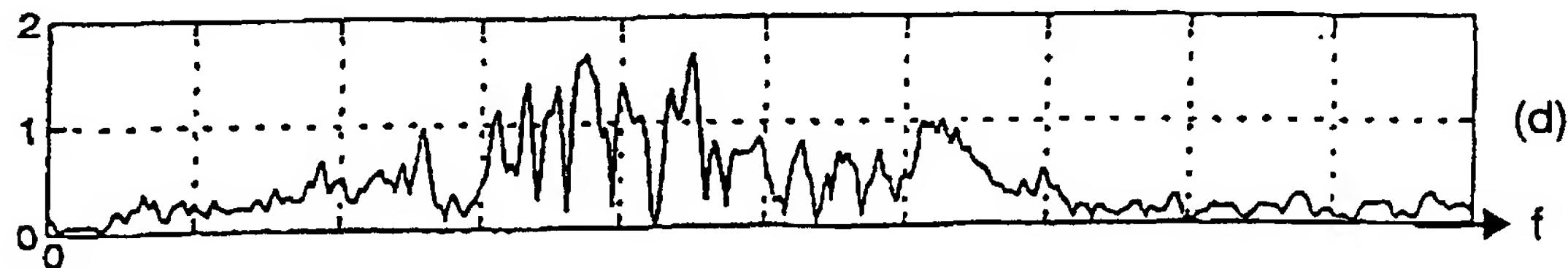
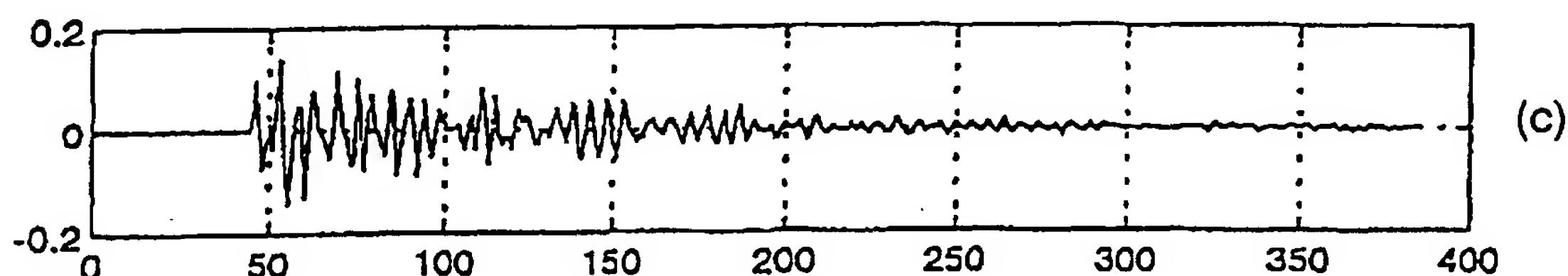
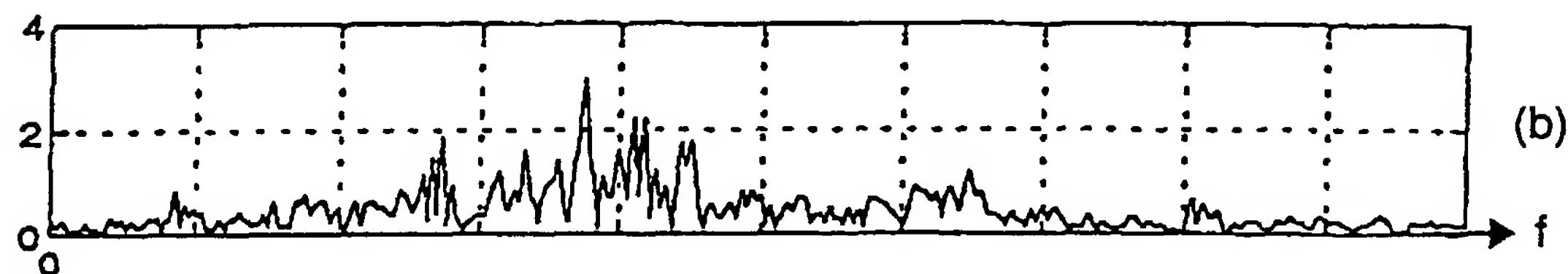
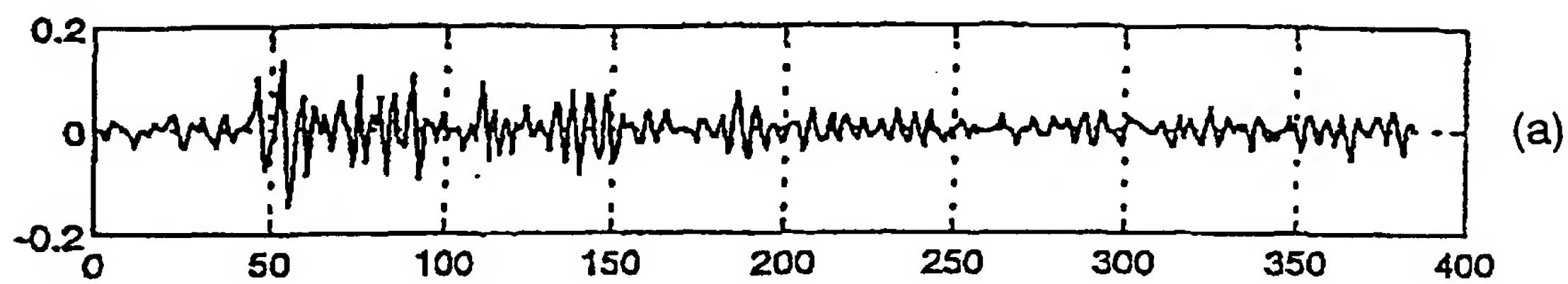


FIG.25

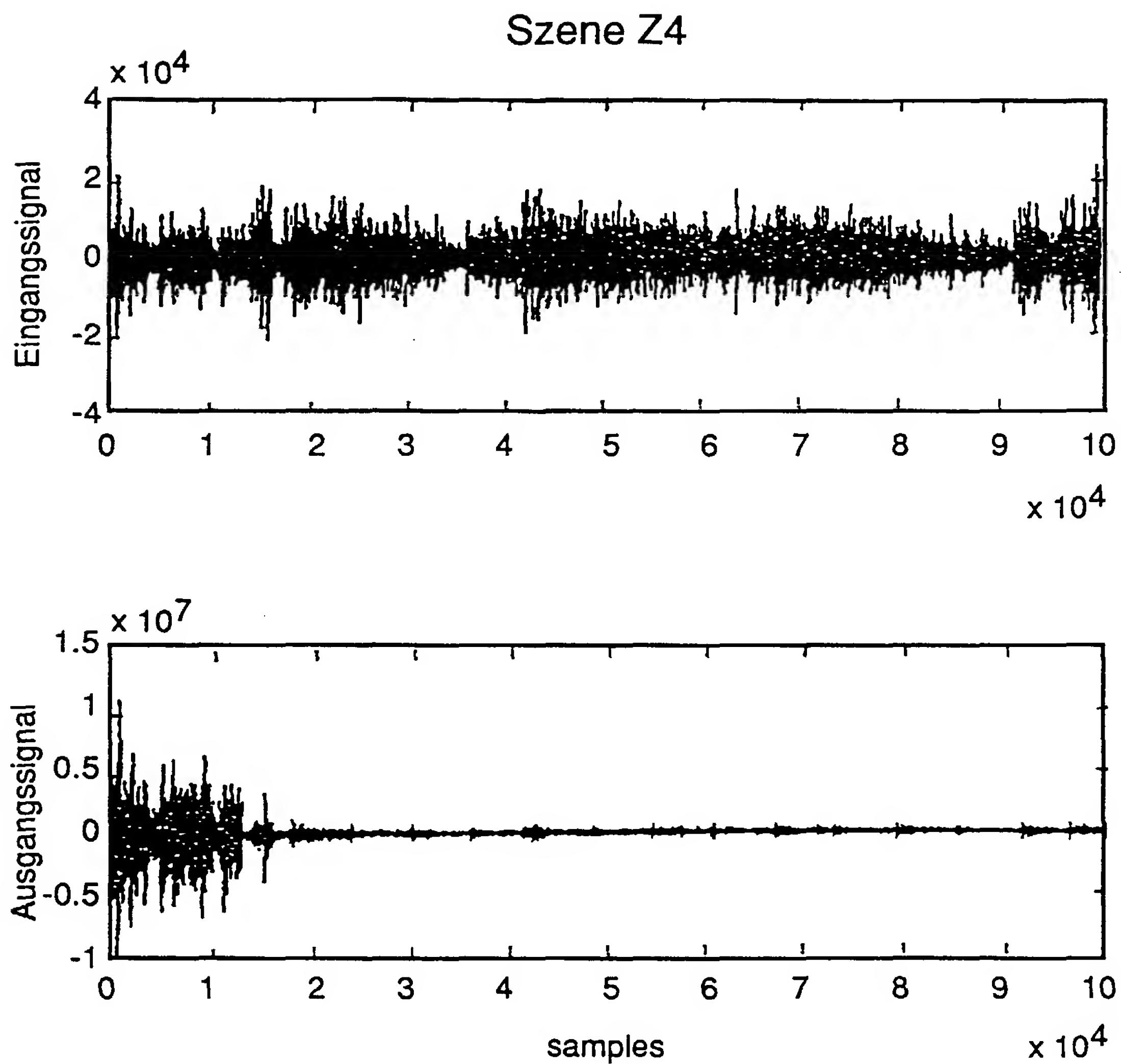
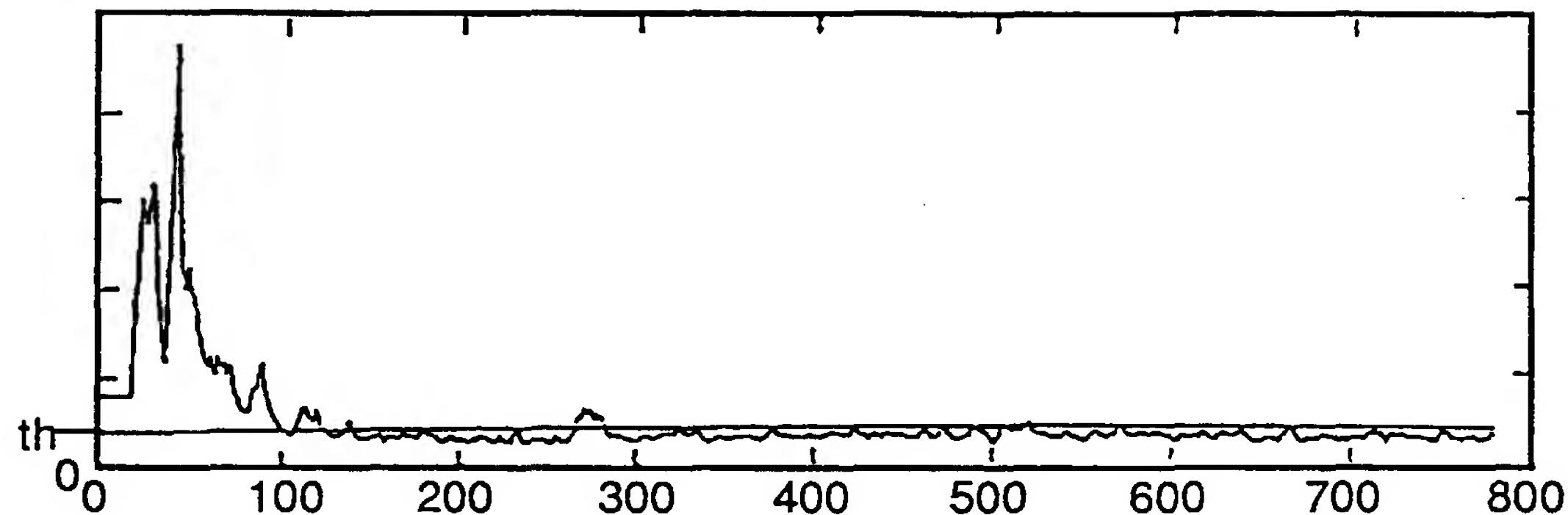


FIG.26

Szene Z4

fea



Speichern

Ein

Aus



FIG.27

Szene Z4

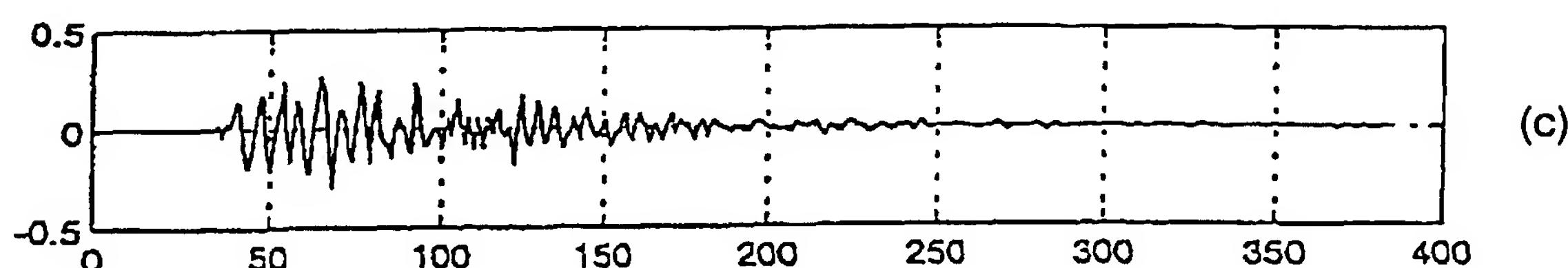
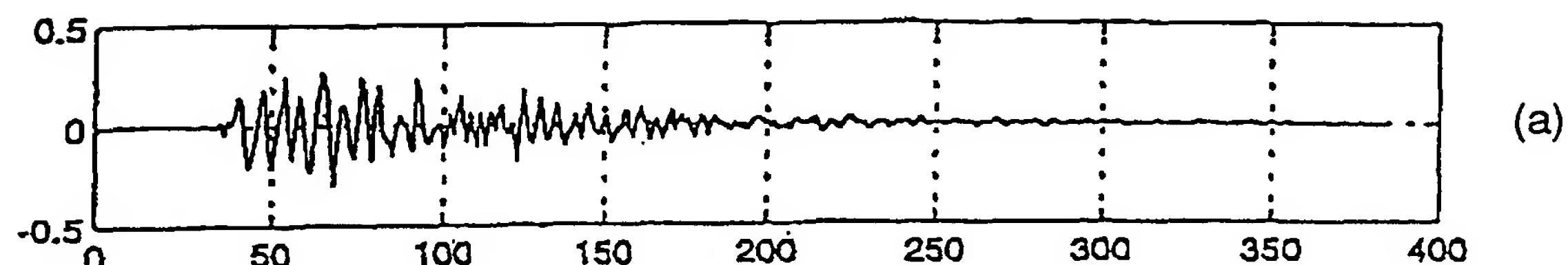


FIG.28